Basque Speecon-like and *Basque SpeechDat MDB-600*: speech databases for the development of ASR technology for Basque

**Igor Odriozola, Inma Hernaez, M. Inés Torres,
Luis J. Rodriguez-Fuentes, Mikel Penagarikano, Eva Navas**

University of the Basque Country (UPV/EHU)
{igor, inma, eva}@aholab.ehu.es, {manes.torres, luisjavier.rodriguez, mikel.penagarikano}@ehu.es

Abstract

This paper introduces two databases specifically designed for the development of ASR technology for the Basque language: the *Basque Speecon-like* database and the *Basque SpeechDat MDB-600* database. The former was recorded in an office environment according to the Speecon specifications, whereas the later was recorded through mobile telephones according to the SpeechDat specifications. Both databases were created under an initiative that the Basque Government started in 2005, a program called ADITU, which aimed at developing speech technologies for Basque. The databases belong to the Basque Government. A comprehensive description of both databases is provided in this work, highlighting the differences with regard to their corresponding standard specifications. The paper also presents several initial experimental results for both databases with the purpose of validating their usefulness for the development of speech recognition technology. Several applications already developed with the *Basque Speecon-like* database are also described. Authors aim to make these databases widely known to the community as well, and foster their use by other groups.

Keywords: Language resources, ASR database, Basque speech database

1. Introduction

Basque is a European minority language with not many speech resources for developing robust speaker independent Large Vocabulary Continuous Speech Recognition (LVCSR) systems. Indeed, according to the study carried out by the META-NET European network of excellence (Hernaez et al., 2012), it exhibits "fragmentary support" at several levels of speech technologies, so Basque is considered among the European languages whose survival is threatened in the digital age.

The only publicly available dataset for the development of speech recognition systems in Basque is *Basque FDB-1060*, which was designed according to the specifications of SpeechDat and recorded over the fixed telephone network, and it is distributed by ELRA (Hernaez et al., 2003). In 2005, the Basque Government started a program, called ADITU, that specifically aimed at developing speech recognition and synthesis technologies for the Basque language. The ADITU program supported the creation of two new speech databases: the first one, that we hereafter call *Basque Speecon-like* database, was recorded in an office environment according to Speecon specifications; the second one, hereafter called *Basque SpeechDat MDB-600* database, was recorded through mobile telephones according to SpeechDat specifications. Both databases belong to the Basque Government.

This paper is organized as follows: in the next two sections, the *Basque Speecon-like* and the *Basque SpeechDat MDB-600* are introduced. Then some applications already developed using the *Basque Speecon-like database* are described. Finally, some conclusions are explained.

2. *Basque Speecon-like database*

The *Basque Speecon-like* speech database was recorded in an office-environment, partially following the specifications of Speecon (Isakra et al., 2002), a shared-cost project funded by the European Commission under Human Language Technologies, which was a part of the Information Society Technologies Programme (IST-1999-10003)[1]. The Speecon project was launched in February 2000 and focused on collecting resources for training speech recognizers. Since the final goal of Speecon was the development of voice-driven interfaces for consumer applications, audio files were recorded via microphone at different distances. Under this framework, databases were produced for 20 European languages, among which Basque was not included.

2.1. Design issues

The *Basque Speecon-like* database was designed with three main types of contents:

a) **Read core words**. This set included task-dependent words related to electronic devices, such as *aktibatu* (activate), *berrasi* (restart), *gelditu* (pause), that were proposed in the Speecon project and translated into Basque. This category also included digit sequences, amounts, times and dates, spellings, etc. Additionally, two lists of frequently used street and city names were created, including those most frequently used by Basque population. Thus the list mainly included Basque places but also Spanish, European and some North American places. In the same way, a set of

[1] http://www.speechdat.org/speecon/index.html

email and frequent web addresses were selected.

b) **Read phonetically-rich sentences**. This set consists of phonetically balanced sentences and words. To define it, a large text corpus consisting of 321 476 sentences that included 5 067 248 running words was considered. This corpus was extracted from both written language, i.e. newspapers and contemporary literature, and spoken language, i.e. transcriptions of speech. Phoneme and phonetic context distributions were estimated and chosen as reference for the Basque language. A set of 9200 phonetically balanced sentences was also defined, which was further divided into 230 subsets of 40 phonetically balanced sentences, each one assigned to a different speaker. To guarantee the phonetic balance in all the 230 subsets, some sentences had to be manipulated to add words including less frequent phonemes. A preliminary list with the 13 000 most frequent words was extracted from the large text corpus, and finally a reduced list of 1841 phonetically rich words (keeping phonetic balance) was then extracted from it.

c) **Spontaneous speech**. This set consists of spontaneous answers to some predefined questions or prompts, as well as a short spontaneous story. Prompts and questions were aimed to get spontaneous answers about dates, times, spellings, companies and people names, city names, phone numbers, language names as well as yes/no answers. Next, six out of 30 possible topics were proposed to the speaker, to elaborate a short story or just give his or her opinion on them. The goal was to get five minutes of free spontaneous speech. The proposed topics included personal interests and hobbies, favourite sports, films, TV series, places that they had visited or would like to visit, etc. Some scenarios were also proposed related to bank transfers, hotel and travel booking, movie tickets, bus routes and schedules, etc.

2.2. Recording platform

The Speecon specifications recommend to record speech signals in four different environments (office, public places, entertainment and car), to cover diverse application scenarios of automatic speech recognition (ASR) technology. However, it was considered that the most interesting applications were those involving office environments, so recordings were made in closed rooms using a desktop computer as recording platform. The recording setup was also simplified, since only two microphone channels were recorded simultaneously: close-talk and desktop, whereas the Speecon standard specifies four: close-talk, lavalier, desktop and far-field. A Shure SM10A was used as close-talk microphone and a Shure SM58 was used as desktop microphone (with a Shure FP11 microphone-to-line amplifier). Audio signals were acquired at 16 kHz with 16-bit linear PCM encoding.

2.3. Distribution of speakers with regard to dialect region and competence level

The distribution of speakers in the database is an issue that must be handled with care, in order to be representative of the community of potential users. The Basque language distribution is very complex with regard to demographics, since it presents many dialectal variations in a very small geographic area and significant variabilities even within the same dialect. The standard Basque is relatively new, to the point that native Basque speakers may not be used to speak standard Basque. That is why most recorded native speakers, though capable of using standard Basque in elicited recordings, use their own dialectal variety in spontaneous speech recordings. The distribution of speakers with regard to dialectal region and competence level is shown in Table 1.

	Native		**Non-native**		**Total**	
	fin.	des.	fin.	des.	fin.	des.
Gipuzkoa	85	71	15	17	100	88
Bizkaia	49	62	47	39	96	101
Nafarroa	14	13	9	5	23	18
Araba	0	6	7	17	7	23
Others	1	-	3	-	4	-
Total	149	152	81	78	230	230

Table 1: Distribution of speakers with regard to dialect and competence level in the *Basque Speecon-like* database. Note the difference between the designed distribution (des.) and the final distribution (fin.).

Due to recruiting issues and schedule requirements, the final distribution of speakers did not exactly match the design. As a result, some regions (Nafarroa and Gipuzkoa) were over-represented, whereas others (especially Araba) were under-represented. In any case, the resulting distribution is still useful and quite representative of the demographics of Basque.

2.4. Distribution of speakers with regard to age and gender

The Speecon specifications considered both adult and children speech recordings. However, the targeted applications of *Basque Speecon-like* database did not require dealing with children speech and only adult speakers were recruited. The total amount of speakers in the *Basque Speecon-like* database is 230 (127 female + 103 male), whereas the amount of speakers in Speecon databases was typically 600. Table 2 summarizes the gender distribution across the age groups defined in the standard.

As can be seen in Table 2, the *Basque Speecon-like* database shows a slight deviation from Speecon specifications with regard to gender distribution. However, the deviation is very small, so we can assume that Speecon specifications are fulfilled.

	Female	Male	Total	%	Speecon specs
15-30	67	38	105	45.65	≥ 30%
31-45	48	51	99	43.04	≥ 30%
46+	11	14	25	10.87	≥ 10%
Unknown	1	0	1	0.44	
Total	127	103	230	100	
%	55.22	44.78	100		
Speecon specs	45-55%	45-55%			

Table 2: Distribution of speakers with regard to age and gender in the *Basque Speecon-like* database.

2.5. Database contents

As explained in section 2.1, the *Basque Speecon-like* database consists of three main types of contents, with two types of speech: one containing read speech, and the other spontaneous speech. The main difference between them is that spontaneous speech is strongly dialectal for native speakers, whereas almost all read speech recordings involve standard Basque (regardless of the competence level). A breakdown of the database contents is shown in Table 3, along with the values of the Speecon standard. Though the Speecon standard involves a higher number of items, most of them are application-specific words, which makes the difference not so relevant. The whole *Basque Speecon-like* database amounts to 23.7 GB.

	Basque Speecon-like	**Speecon**
Spontaneous speech		
Free spontaneous items	5	10
Elicited answers	18	17
Read speech		
Phonetically rich sentences	40	30
Phonetically rich words	8	5
General purpose words & phrases	32	31
Application specific words & phrases	212	453

Table 3: *Basque Speecon-like* database contents (number of utterances per speaker) and comparison to the Speecon standard.

2.6. Annotation

All the recordings were made at the time the ADITU program was active, but some of the annotations, including an improved lexicon and improved transcriptions of the spontaneous speech recordings, have been produced later, in an effort to have more reliable and useful datasets for the development of ASR systems in Basque.

2.6.1. Orthographic annotation procedure

In the case of read speech, the orthographic transcriptions of the audio files were available beforehand. Thus, these transcriptions were just checked and corrected when needed. In the case of spontaneous speech, the whole transcriptions were manually created from scratch and later checked to fix errors and to increase consistency.

2.6.2. Acoustic events

A set of labels was used to mark acoustic events and word deformations (see Table 4). This allowed us to keep as much speech as possible in the database, avoiding the need to take out recordings from the corpus.

	Symbol	Meaning
speech events	{FIL}	Filled pause
	{FRA}	Word fragment
	{LNT}	Lengthened word
	{TRC}	Truncated word
	{UNI}	Unintelligible word
non-speech events	{BRE}	Breath (or laugh)
	{INT}	Intermittent noise
	{SPK}	Speaker noise (lips, smacks etc.)
	{STA}	Stationary background noise

Table 4: Labels used to mark acoustic events and word deformations in the *Basque Speecon-like* database.

2.6.3. Phonetic transcription

An improved lexicon has been created, containing all the orthographic entries together with their broad phonetic transcriptions according to the SAMPA Basque alphabet[2]. A basic set of 36 phones was considered (35 phones belonging to the standard inventory and 1 dialectal phone). The improved lexicon has been automatically generated using a grapheme-to-phoneme (G2P) transcriber for Basque, including as alternatives the different pronunciations that the same standard word shows in different dialects. The lexicon currently contains 29 780 different lexical entries (in standard Basque), with 131 243 different pronunciations. This means that each word has on average 4.41 different pronunciations, which typically account for dialectal variations, revealing the complexity of Basque from the point of view of its use in the daily life.

[2] http://aholab.ehu.es/sampa_basque.htm

2.7. Database validation

Several speech recognition experiments have been carried out in order to validate the usefulness of the *Basque Speecon-like* database. Here we briefly describe some of them:

a) **Isolated word recognition experiments**: An internal validation of the *Basque Speecon-like* database was performed through an isolated-word recognition experiment (Odriozola, 2007). Two experiments were carried out: one for the recognition of application specific words (read speech), and another one for the recognition of proper names and city names (elicited spontaneous speech). A closed lexicon of 435 words was used for the first experiment whereas a lexicon of 19 383 words was used for the second one. The acoustic models were estimated using Mel-Filter Cepstral Coefficients (MFCC) as acoustic parameters and continuous hidden Markov models (HMM) with 16 Gaussian mixtures as acoustic models. HTK (Young et al., 2006) was used to train the models as well as to perform the decoding process. During the development of the experiments, corrections to the initial lexicon were done together with some adaptations in the G2P transcriber to cope with dialectal speech. Results are shown in Table 5.

	Application specific words	**Proper names and city names**
Speech type	Read speech	Elicited speech
Nº files tested	11 859	584
Lexicon size	435	19 383
Recognition accuracy (%)	98.85	89.04

Table 5: Experimental conditions and results of isolated-word recognition experiments on the *Basque Speecon-like* database.

b) **Continuous speech recognition experiments**: Though not yet published, another set of both isolated-word and continuous speech recognition experiments has been carried out on different subsets of the *Basque Speecon-like* database, using continuous HMMs with 32 Gaussian mixtures as acoustic models. The models were initialized using a small phonetic database in Basque and then retrained on the phonetically balanced subset of sentences of the *Basque Speecon-like* database. Different speech recognition experiments were performed on different subsets: isolated digits, digit sequences, application specific words and free speech. The evaluation was performed both at word level and at phone level. Table 6 shows the experimental conditions and the results of those experiments.

	Digits	**Digit sequences**	**Application specific words**	**Free speech**
Speech type	Read	Read	Read	Spontaneous
Nº files tested	1150	230	48 300	1380
Lexicon size	12	12	609	15 407
Phone acc. (%)	55.14	46.71	63.80	53.42
Phone accuracy excluding insertion errors (%)	83.00	72.80	81.73	60.47
Word acc. (%)	87.20	92.19	83.29	--

Table 6: Experimental conditions and results of speech recognition experiments for different subsets of the *Basque Speecon-like* database.

3. *Basque SpeechDat MDB-600*

The *Basque SpeechDat MDB-600* database was designed following the guidelines of the mobile section[3] of the SpeechDat project (Elenius et al., 1997). The SpeechDat project was a CEC-funded initiative (LRE-63314) that addressed the fields of production, standardization, evaluation and dissemination of Spoken Language Resources (SLR). One of its main tasks was to help to create a European infrastructure for distribution and evaluation of SLR.

3.1. Recording platform

Fulfilling SpeechDat requirements, a PC-based recording system with an ISDN interface board connected to a local exchange was set. Signal files were recorded using a Windows XP based system using an AVM-ISDN board. Files were stored directly onto hard disk and backed up regularly. One PC was used giving a maximum capacity of 2 parallel calls. The database was recorded over a 4 months period where 951 call attempts were received. Only 602 of these calls were completed, recorded and accepted. The maximum load on a single day was 25 calls. The speech files were stored as sequences of 8 bit - 8 kHz A-law samples. Each prompted utterance was stored within a separate file and the associated label files were stored in SAM file format. One of the characteristics of a SpeechDat database recorded over a mobile network is that the calls must be collected in a balanced way from 4 different environments (office, public place, place with background noise and inside a moving vehicle) with a maximum deviation of 5%. The final distribution of calls also complies with this requirement.

[3] http://www.speechdat.org/speechdt/speechdat_m/home.html

3.2. Distribution of speakers with regard to dialectal and competence level

The issue of the selection of speakers with regard to their dialect region and competence level was solved in the same way as for the *Basque Speecon-like* database (see section 2.2).

3.3. Distribution of speakers with regard to age and gender

Table 7 shows the distribution of speakers in the database, in terms of gender and age, as well as those specified by the standard. As can be seen in the table, the *Basque SpeechDat MDB-600* database fits the SpeechDat mobile requirements.

	Female	Male	Total	%	SpeechDat specs
Under 16	2	2	4	0.66	
16-30	144	137	281	46.68	≥ 20%
31-45	97	88	185	30.73	≥ 20%
46-60	67	51	118	19.60	≥ 15%
Over 60	5	9	14	2.33	
Unknown	0	0	0	0,00	
Total	315	287	602	100	
%	52.33	47.67	100		
SpeechDat specs	45-55%	45-55%			

Table 7: Distribution of speakers with regard to age and gender in the *Basque SpeechDat MDB-600* database.

	Basque SpeechDat MDB-600	**SpeechDat**
Spontaneous speech	**10**	**10**
Free spontaneous items	1	-
Elicited answers	9	-
Read speech	**44**	**44**
Phonetically rich sentences	9	-
Phonetically rich words	4	-
Gen. purpose words & phrases	22	-
Application specific words and phrases	9	-

Table 8: *Basque SpeechDat MDB-600* database contents (number of utterances per speaker) and comparison to the SpeechDat standard.

3.4. Database contents

The general specifications for the SpeechDat mobile speech database recommend the recording of a minimum of 500 speakers, 54 items each call (10 corresponding to spontaneous speech and 44 to read speech). The final specification for the *Basque SpeechDat MDB-600* recordings is shown in Table 8. The whole *Basque SpeechDat MDB-600* database amounts to 1.70 GB.

3.5. Annotation

3.5.1. Orthographic annotation procedure

Only complete recordings were transcribed (recordings for which signal files exist for every prompt) and prompts were shown to transcribers so that they could concentrate on unpredictable phenomena. A single transcriber was in charge of the labelling of the database.

3.5.2. Acoustic events

Symbols used to denote truncations, mispronunciations, unintelligible speech and non-speech acoustic events are in accordance to SD1.3.2 (Senia et al., 1997). A summary can be seen in Table 9. In addition, the "+" sign is used to denote coarticulation between words, a typical Basque language phenomenon which consists on losing the pronunciation of the end of some words when they appear in combination with some other words. Besides, the symbol "." was used to mark very long silences in the recordings.

	Symbol	Meaning
speech events	*abc	Mispronunciations
	**	Unintelligible speech
	&abc	GSM distortion
non-speech events	{FIL}	Filled pause
	{INT}	Intermittent noise
	{SPK}	Speaker noise (lips, smacks etc.)
	{STA}	Stationary background noise

Table 9: Labels used to mark acoustic events and word deformations in the *Basque SpeechDat MDB-600* database.

3.5.3. Phonetic transcription

A list of phonetic transcriptions for every word occurring in the orthographic transcription is provided along with the audio files. The spelling of each entry in the lexicon was automatically checked. Some words are pronounced differently due to dialectal differences, so several alternatives are also taken into account. The phonetic transcriptions are described with 34 SAMPA Basque phoneme symbols (33 phones belonging to the standard inventory, omitting the semivowels, and 1 dialectal phone). The word pronunciations were generated using a G2P transcriber.

3.6. Database validation

The experiments carried out to validate the database are those specified in the *RefRec* (the COST 249 SpeechDat Reference Recogniser) procedure (Johansen et al., 2000), which aimed at setting reference results for any language developing speech recognition technology. The tests were done over different subcorpora representing typical test applications:

I: Isolated digits
Q: Yes/no
A: Isolated application words
BC: Connected digit strings, unknown length
O: City names
W: Phonetically rich words

The acoustic features were conventional 39-dimensional MFCCs, including zero'th cepstral coefficient C_0, as well as first and second order deltas. An extra experiment using normalized CMVN cepstra (Viiki et al., 1998) was also carried out in order to see how the results could improve by means of partially removing the effects of different communication channels and background noises. The results of the experiments are shown in Table 10, where the Word Error Rates obtained for the *Basque SpeechDat MDB-600* are shown together with the results of the *Swedish MDB-1000* for comparison. This language has been chosen because it is the only one that presents results for mobile devices in the previously mentioned *RefRec* paper (Johansen et al., 2000). Table 10 also shows the dictionary size for each test.

Test corpus	*Basque MDB-600*			*Swedish MDB-1000*	
	Dict. size	WER %	WER % (CMVN)	Dict. size	WER %
I	12	12.73	3.25	10	10.5
Q	2	7.47	1.18	2	1.13
A	42	18.14	7.18	30	4.04
BC	12	4.06	2.58	10	14.22
O	1352	38.53	20.79	869	18.59
W	1055	37.65	18.76	3611	52.35

Table 10: Dictionary sizes and WER of speech recognition experiments carried out on different test subsets (Refrec), for *Basque SpeechDat MDB-600* and *Swedish MDB-1000*.

4. Applications

The *Basque Speecon-like* database has already been used by the research groups of the University of the Basque Country involved in this paper. In this section, we describe some applications that have been developed recently:

a) **Speech recognition experiments using an external LM:** The main objective of this experiment was to perform a continuous speech recognition test by using the acoustic models obtained from the *Basque Speecon-like* database along with an external language model (LM). Firstly, the database was divided into a training set (155 sessions) and a test set (75 sessions). The acoustic models were trained using the subset of read speech from the training set, using MFCCs as acoustic features and HMMs as models. The LM was created from the CRP (Contemporary Reference Prose), a textual corpus collected by the Basque Institute of the University of the Basque Country, which contains some 25.1 million words. 13.1 million of these words are drawn from books chosen for their quality (287 volumes) and 12 million come from newspaper articles. Thus, a word-level model was created composed of 984 238 unigrams, 12 558 022 bigrams and 3 004 799 trigrams.
The speech files tested in the experiment were those belonging to the subset of phonetically rich sentences of the test set (2950 sentences), using all the vocabulary of the training set (23 243 words) and the external LM. A coverage analysis showed that the test set contained 40.71 % of out-of-vocabulary (OOV) words, which was expected due to the agglutinative nature of Basque. So a second experiment was carried out adding those words to the vocabulary (26 624 in total). In both experiments a set of 31 phonetic units was used. Another experiment was carried out after joining some phonetic units, thus using a set of 24 units and removing pronunciation alternatives from the training dictionary. Normalized CMVN cepstra were used as explained in (Viiki et al., 1998). The results are shown in Table 11.
These experiments were performed using the HMM-based recognizer developed in the Aholab research group of the University of the Basque Country.

	Dict. size	Phon. units	Word acc. (%)
Exp. 1	23 243	31	69.38
Exp. 2	26 624	31	79.56
Exp. 3	26 624	24	80.53

Table 11: Dictionary size, number of phonetic units and Word Accuracy of three experiments on the phonetically rich sentences subset of the *Basque Speecon-like* database.

b) **Speech recognition for a spoken document retrieval system**: A spoken document retrieval system was developed, based on the automatic transcription of speech contents, which retrieved audio/video segments from a TV news repository (Varona et al., 2011). To evaluate the performance of the system for

Basque, 7 manually transcribed TV broadcasts in Basque were used as test set. The speech recognizer was trained on the *Basque Speecon-like* database, using MFCCs as acoustic features and continuous HMMs as acoustic models, with 64 Gaussian mixtures and a simplified set of 26 phonetic units. A subset of 5346 read sentences from 140 speakers was used to train the acoustic models. The language model was trained on text news in Basque taken from the internet from 2003 to 2008, including 2 589 284 sentences, 34 510 770 words and a vocabulary of 661 651 words. Three different LMs were trained, based on two reduced vocabularies containing the 5000 and 20 000 most likely words in the text news, and a closed vocabulary containing just the 8903 words appearing in the Basque TV broadcasts used as test set. Experiments were carried out on three sets of test segments: (1) presenter (planned speech in studio conditions); (2) reporter (planned but less formal speech recorded outdoor, probably with background noise); and (3) spontaneous (interviews, commonly including disfluencies and background noise). Table 12 shows the number of segments, the number of words, the vocabulary size and the results obtained for each of the test sets.

	Test set size			**Word accuracy (%)**		
	#segm	#wrd	#voc	5K	20K	Closed
presenter	180	2971	1707	49.85	59.81	91.89
reporter	180	2885	1626	49.57	60.67	92.65
spontaneous	90	1569	807	27.68	28.51	60.97

Table 12: Word-level speech recognition results on different types of broadcast news speech in Basque. Details about the size of the datasets are also provided. In all cases, the *Basque Speecon-like* database was used to train the acoustic models.

c) **Development of a Computer-Assisted Pronunciation Training (CAPT) system for Basque** (Odriozola et al., 2012): A method to build CAPT systems for low-resourced languages such as Basque is proposed. Due to the lack of specific acoustic resources, the main goal of the work was to create a pronunciation evaluation system using a general purpose ASR speech database such as the *Basque Speecon-like* database. The method, which was based on continuous HMMs, automatically determined the thresholds of Goodness of Pronunciation (GOP) scores. For that purpose, two GOP distributions were obtained for each phone: on the one hand, the GOP distribution when the phone is correctly pronounced; on the other hand, the GOP distribution when the phone is incorrectly pronounced. The latter was obtained simulating errors and evaluating them in forced alignment mode. Then the threshold corresponding to each phone was calculated using the Equal Error Rate (EER) parameter. Results showed that the performance of the system differs noticeably depending on the phone being evaluated. The discernment of the phones that do not appear in Spanish is not as accurate as the ones' that belong to both phone inventories. This issue can be partially solved by removing the speech of non-native speakers from the training set. The method proves to be very useful when there is no database specifically designed for CAPT systems, although it is not as accurate as those specifically designed for this task.

5. Conclusions

In this paper, we have described two speech databases for the development of ASR technology for Basque: the *Basque Speecon-like* database and the *Basque SpeechDat MDB-600* database, both belonging to the Basque Government. Besides showing the characteristics of both databases in detail, focusing on the differences with regard to the respective standards, the results of several speech recognition experiments have been also presented, in an effort to validate the usefulness of the databases.

6. Acknowledgements

This work has been partially supported by the Spanish Ministry of Economy and Competitiveness (SpeechTech4All project, TEC2012-38939-C03-03), the Basque Government (Ber2tek project, IE12-333; Saiotek project, S-PE13UN105) and the University of the Basque Country (grant GIU13/28).

7. References

Elenius, K., & Lindberg, J. (1997). SpeechDat - Speech databases for creation of voice driven teleservices. In Bannert, R., Heldner, M., Sullivan, K., & Wretling, P. (Eds.), *Proceedings of Fonetik - 97*, Dept of Phonetics, Phonum 4, pp. 61-64.

Hernaez, I., Luengo, I., Navas, E., Zubizarreta, M., Gaminde, I., Sanchez, J. (2003). The Basque speech_dat (II) database: a description and first test recognition results, In *Eurospeech-2003*, pp. 1549-1552.

Hernaez, I., Navas, E., Odriozola, I., Sarasola, K., Diaz de Ilarraza, A., Leturia, I, Diaz de Lezana, A., Oihartzabal, B., Salaberria, J. (2012). *The Basque Language in the Digital Age (White Paper Series, Georg Rehm, Hans Uszkoreit, Series Eds)*, Springer, Heidelberg, New York, Dordrecht, London, September 2012. URL http://www.meta-net.eu/whitepapers/volumes/basque.

Isakra, D., Großkopf, B., Marasek, K, Van den Heuvel, H., Diehl, F., Kiessling, A. (2002). SPEECON – Speech Databases for Consumer Devices: Database Specification and Validation. *Proc. of LREC 2002*, pp. 329-333.

Johansen, F. T., Warakagoda, N., Lindberg, B., Lehtinen, G., Kacic, Z., Zgan, A., Elenius, K., Salvi, G. (2000).

COST 249 SpeechDat Multilingual Reference Recogniser. Proc. of LREC 2000, Athens, Greece.

Odriozola, I. (2007). Development of an isolated word recognition system for Basque. Master thesis (in Basque). Available in: http://ixa.si.ehu.es/master/master_tesiak/1305202902/publikoak/Master_tesia_Igor_Odriozola.pdf

Odriozola, I., Navas, E., Hernaez, I., Sainz, I., Saratxaga, I., Sánchez, J., Erro, D. (2012). Using an ASR database to design a pronunciation evaluation system in Basque. *Proc. of LREC 2012*, pp. 4122-4126.

Senia, F., Comeyne, R., Lindberg, B. (1997). *Enviromental and speaker specific coverage for Fixed Networks*, SpeechDat Deliverable LE-4001-SD1.3.2 version 2.2.

Varona, A., Nieto, S., Rodriguez-Fuentes, L.J., Penagarikano, M., Bordel, G., Diez, M. (2011). A Spoken Document Retrieval System for TV Broadcast News in Spanish and Basque. *Procesamiento de Lenguaje Natural, Vol. 47*, pp. 77-85.

Viiki, O., Laurila, K. (1998). Cepstral domain segmental feature vector normalization for noise robust speech recognition. *Speech Communication 25*, pp. 133-147.

Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., Woodland, P. (2006). *The HTK Book Version 3.4*. Cambridge University Engineering Department (UK).