

The Alborada-I3A Corpus of Disordered Speech

Oscar Saz, Eduardo Lleida, Carlos Vaquero, W.-Ricardo Rodríguez

Communications Technology Group (GTC)
Aragón Institute for Engineering Research (I3A)
University of Zaragoza, Spain
{oskarsaz,lleida,cvaquero,wricardo}@unizar.es

Abstract

This paper describes the “Alborada-I3A” corpus of disordered speech, acquired during the recent years for the research in different speech technologies for the handicapped like Automatic Speech Recognition (ASR) or pronunciation assessment. It contains more than 2 hours of speech from 14 young impaired speakers and nearly 9 hours from 232 unimpaired age-matched peers whose collaboration was possible by the joint work with different educational and assistive institutions. Furthermore, some extra resources are provided with the corpus, including the results of a perceptual human-based labeling of the lexical mispronunciations made by the impaired speakers. The corpus has been used to achieve results in different tasks like analyses on the speech production in impaired children, acoustic and lexical adaptation for ASR and studies on the speech proficiency of the impaired speakers. Finally, the full corpus is freely available for the research community with the only restrictions of maintaining all its data and resources for research purposes only and keeping the privacy of the speakers and their speech data.

1. Introduction

Research in Human Language Technologies (HLT) for the handicapped is a field of growing interest within the community. A growing number of oral-based control systems are being currently worked on to provide of assistive technology to handicapped individuals and improve their quality of life using Automatic Speech Recognition (ASR) (Parker et al., 2006) or Text-to-Speech (TTS) synthesis (Creer et al., 2009). Furthermore, novel tools for Computer-Aided Speech and Language Therapy (CASLT) can help reduce the communicative disabilities of those affected with physical and cognitive impairments (García-Gómez et al., 1999; Vicsi et al., 1999; Hatzis et al., 2003). A major challenge in this task is the sparse existence of corpora which can provide of well-established benchmarks to be used within the community. Several corpora covering different aspects of disordered speech have been released through the years like the Whitaker database (Deller et al., 1993), the Nemours database (Menéndez-Pidal et al., 1996) or, more recently, the Universal Access Database (Kim et al., 2008). Each one of these corpora presents a different set of features in terms of age or type of impairment of the recorded individuals, type of language acquired or presence of multiple sources of information in multimodality. This sparsity of resources is more significant in languages like Spanish, where very little corpora like the HACRO project corpus (Navarro-Mesa et al., 2005) had faced previously this specific area of research.

The motivation of this work arose, hence, to fill the gap in this area in Spanish while trying to provide novel resources to the whole community in an attempt to attract the attention towards this line of research. With this starting point, the authors from the Aragón Institute for Engineering Research (I3A) initiated a collaborative work with the staff of the Public School for Special Education (CPEE) “Alborada” from Zaragoza (Spain) oriented to the acquisition of a speech corpus containing disordered speech.

The requirements of the corpus were to acquire speech from

several speakers covering a wide range of disorders capturing all their acoustic and lexical special properties while assuring the best ratings of quality in the speech acquisition process in terms of noise and environmental distortion. The corpus has also been including new features along time like the outcome of a human labeling on the speakers mispronunciations. With all these features, it was considered that it was the moment to make it public for the community to share and use, incrementing the awareness on the use of HLTs for the improvement of the quality of life of the handicapped.

This paper is organized as follows: Section 2. will describe the speakers and sessions of the originally recorded corpus with a series of young impaired speakers. Section 3. will present the works carried out around the original acquisition to complete and expand the possibilities of the corpus. Later on, Section 4. will introduce some research results achieved with the corpus. Finally, Section 5. will provide the conclusions of the work, in terms of applications of the corpus and availability for further research.

2. Description of the Original Corpus

As mentioned, the initial corpus acquisition was planned to only involve the recordings from a small set of speakers with speech impairments from the CPEE “Alborada”. This corpus and the acquisition process which was undertaken can be described through 3 features: Recording environment, selection of speakers, and design of the recording sessions.

2.1. Recording environment

The selection of the recording environment was made to fulfill two objectives: Provide a low-noise environment while assuring that the speakers felt confident and comfortable, because they mostly were children with cognitive disabilities. Inserting these speakers in a specific recording environment like an anechoic or specially designed room could put them in a situation of stress which might have

Speaker	Gender	Age	Disorders	Speaker	Gender	Age	Disorders
Spk001	F	14yo	Down's Syndrome	Spk002	M	11yo	Hyperactivity
Spk003	M	21yo	Deprivation disorder	Spk004	F	21yo	Deprivation disorder
Spk005	M	18yo	Down's Syndrome	Spk006	M	17yo	Motor ataxia, tetraplegia
Spk007	M	18yo	Polymalformations	Spk008	M	19yo	Cerebral palsy
Spk009	F	11yo	Development disorder	Spk010	F	15yo	Encephalopatya
Spk011	F	20yo	Cognitive disorder	Spk012	M	18yo	Expressive disability
Spk013	F	13yo	Down's Syndrome	Spk014	F	11yo	Development disorder

Table 1: Speakers in the “Alborada-I3A” corpus

lead to a blockage in their language production. For this reason, the recordings were made in the same facilities of the CPEE “Alborada” where the speakers attended their regular classes to assure the comfort of the speakers and the maximum naturalness in their speech. A member of the I3A supervised the recording process, sometimes accompanied by a staff member of the CPEE “Alborada”. Full supervision of the process assured a fast and reliable acquisition procedure, as cognitive impairments in the speakers might difficult their capabilities to follow up the whole process.

The speech acquisition hardware was a commercial laptop with an integrated soundcard, digitizing the input speech signal with 16 kHz as sampling frequency and 16 bits of depth. These features complied with the usual acquisition characteristics in most of the speech research databases used for similar purposes (ASR or verification). To achieve a low noise level in the recordings, a close-talk microphone (model AKG C444L) was used to reduce the effect of the environment noises. Assuring this quality with the headset microphone was required after selecting an educative center like the CPEE “Alborada” for the recordings, where a certain level of noisiness could appear in the surrounding classes. Anyways, during all the process the speaker and the supervisor(s) were alone in a room the whole time during the recordings. At the end of the recordings, the average Signal-to-Noise Ratio (SNR) across all the utterances was measured in 34.58 dB, indicating the final quality of the recordings.

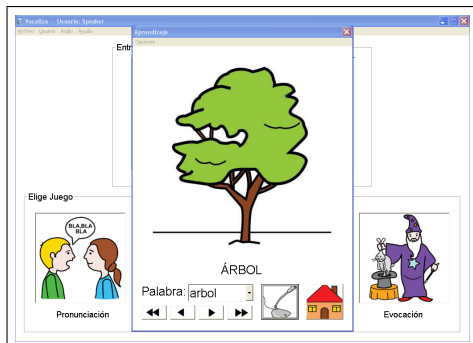


Figure 1: “Vocaliza” acquisition window

The recording process was carried out throughout the Computer-Assisted Pronunciation Training (CAPT) tool “Vocaliza” (Vaquero et al., 2008), part of the set of CASLT tools developed in “Comunica” (Saz et al., 2009d) with

the objective of providing a computer-based framework for speech therapy in Spanish in all different stages of language acquisition. As shown in Figure 1, this tool made use of text, audio and pictorial resources to prompt a certain word or sentence to the user and capture it. Moreover, “Vocaliza” allowed the supervisors to check the quality of the resulting speech signal priorly to store it within the corpus or discard it and repeat the recording of the same word.

2.2. Speaker selection

The process of selection of the speakers was initiated by request from the I3A to the CPEE “Alborada” staff, according to the possibilities of their pupils. The intention was to count on a set of speakers balanced in terms of gender and age withing the range of 11 to 21 years old, as this is the maximum age in which students can stay in a public institution like the CPEE “Alborada”. Furthermore, the distinct cognitive abilities of each student limited their possibilities as potential speakers in the corpus, as some of them were not able to follow the structured process of speech recordings in a reliable way. Finally, 14 speakers were selected, whose description can be seen in Table 1.

All the speakers suffered from different physical, development and cognitive disabilities which produced several speech and language disorders in them, including dyslalia, dysarthria, Specific Language Impairment (SLI) and other semantic, syntax and pragmatic disorders. The diversity of characteristics in the speakers assured a correct representation of a wide variety of speech disorders in such a limited number of speakers.

2.3. Session design

Three types of sessions were designed to be fulfilled by the speakers in the corpus: Isolated words, simple meaningless sentences and complex meaningful sentences. With this design of the sessions, it was expected to cover growing language abilities from the purely phonological to the high linguistic capabilities.

The basis of the corpus were the isolated word recordings, where each speaker had to utter 4 sessions of the set of 57 speech therapy words included in the Induced Phonological Register (RFI) (Monfort and Juárez-Sánchez, 1989), a very well known handbook for speech therapy in Spanish language. The orthographic prompts for these words, which had also been used in the HACRO corpus (Navarro-Mesa et al., 2005), are presented in Table 2. Each session was recorded in a different day to reflect intra-speaker variability. As a result, the total amount of data acquired during the

Prompt	SAMPA	Prompt	SAMPA	Prompt	SAMPA	Prompt	SAMPA
árbol	[ˈArBOl]	boca	[ˈˈboka]	bruja	[ˈˈbruxa]	cabra	[ˈˈkaBra]
campana	[kAmˈˈpanaː]	caramelo	[karaˈˈmelo]	casa	[ˈˈkasa]	clavo	[ˈˈklaBo]
cuchara	[kuˈˈtSara]	dedo	[ˈˈdeDo]	ducha	[ˈˈdutSa]	escoba	[esˈˈkoBa]
flan	[ˈˈflAn]	fresa	[ˈˈfresa]	fuma	[ˈˈfuma]	gafas	[ˈˈgafAs]
globo	[ˈˈgloBo]	gorro	[ˈˈgorro]	grifo	[ˈˈgrifo]	indio	[ˈˈindjo]
jarra	[ˈˈxarra]	jaula	[ˈˈxawla]	lápiz	[ˈˈlapIT]	lavadora	[laBaˈˈDora]
luna	[ˈˈluna]	llave	[ˈˈLaBe]	mariposa	[mariˈˈposa]	moto	[ˈˈmoto]
niño	[ˈˈniːJo]	ojo	[ˈˈOxo]	pala	[ˈˈpala]	palmera	[pAlˈˈmera]
pan	[ˈˈpAn]	peine	[ˈˈpEjne]	periódico	[peˈˈrjoDiko]	pez	[ˈˈpeT]
piano	[ˈˈpjano]	pie	[ˈˈpje]	piña	[ˈˈpiJa]	pistola	[pIsˈˈtola]
plátano	[plaˈˈtanoː]	playa	[ˈˈplajja]	preso	[ˈˈpreso]	pueblo	[ˈˈpueBlo]
puerta	[ˈˈpwerta]	ratón	[raˈˈton]	semáforo	[seˈˈmaforo]	silla	[ˈˈsiLa]
sol	[ˈˈsOl]	tambor	[tAmˈˈBOr]	taza	[ˈˈtaTa]	teléfono	[teˈˈlefono]
toalla	[toˈˈaLa]	toro	[ˈˈtoro]	tortuga	[tOrˈˈtuGa]	tren	[ˈˈtren]
zapato	[Taˈˈpato]						

Table 2: Words in the RFI (orthographic prompts and SAMPA transcription)

whole process was 3192 utterances in 2h17m4s of speech including silence.

For the acquisition of connected speech, a set of sessions similar to those in the Nemours database (Menéndez-Pidal et al., 1996) were designed. First, a set of simple meaningless sentences were created from the RFI words with the following structure:

el/la (the) RFIWord1 y (and) el/la (the) RFIWord2

Four sessions were designed, each one containing 28 different sentences from this set. Each word appeared twice in each session, one in the position *RFIWord1* and the other one in the position *RFIWord2* to balance the presence of all the words in the sentences. Due to the severe cognitive limitations of the speakers, only *Spk001*, *Spk004*, *Spk006* and *Spk011* had sufficient capabilities to fulfill this task, leading to the presence of only 448 utterances and 25m30s of data in the whole corpus.

To complete the acquisition of connected speech, ten complex meaningful sentences were created, each one with 3 RFI words and a maximum of 9 words in total. Only speakers *Spk004*, *Spk006* and *Spk011* could fulfill this task, leading to 30 utterances in 2m9s of data.

3. Extensions to the Corpus

Once assured the quality of the original corpus, it was decided to provide it with more functionalities and create a whole independent resource around it. Three extra tasks were performed to provide this enhanced usability in research: The expanded acquisition of more sessions from two of the initial speakers, the acquisition of a reference corpus with unimpaired age-matched children; and the labeling of the lexical mispronunciations within the corpus.

3.1. Expansion of speakers *Spk007* and *Spk008*

Two years after the original recordings, 4 more isolated word sessions were recorded from *Spk007* and *Spk008*. These extra sessions appear in the corpus as speakers *Spk107* and *Spk108* to distinguish the original and the later

recordings. The idea behind this extra acquisition was to provide a larger amount of data for a reduced set of the speakers with the aim of evaluation adaptation algorithms which may require bigger amounts of speech. As these two speakers were 18 and 19 years old during the initial acquisition process (and, hence, 20 and 21 in the second round of recordings) it can be expected that no major changes occurred in their speech due to natural reasons like vocal tract modifications or growing, which could have been present in younger speakers.

Age	Males	Females	Age	Males	Females
10yo	15	16	11yo	15	16
12yo	15	15	13yo	15	23
14yo	11	21	15yo	11	11
16yo	15	9	17yo	14	10
Total	111	121			

Table 3: Speakers in the reference corpus

3.2. Reference sub-corpus

HLTs usually require of well-matched data to provide the best performance. Task and domain adaptation are fine solutions to provide enhanced results in several tasks like ASR. The age of the speakers in the “Alborada-I3A” corresponded to child and young speech features (Lee et al., 1999) which are not usually covered by standard corpora.

For this reason, a reference sub-corpus from speakers in the same age range was acquired to provide of this task and domain adaptation. A single isolated RFI words session was designed for each speaker, with the same acquisition process than for the impaired speakers in the corpus depicted in Section 2.1.. Recordings were made in the facilities of the Public School for Primary Education (CEIP) “Río Ebro” and in the Secondary Education Institutes (IES) “Tiempos Modernos” and “Félix de Azara”, where all the students attended their regular classes.

The number of speakers for each age is presented in Table

Speaker	Correct	Substituted	Deleted	Speaker	Correct	Substituted	Deleted
Spk001	98.88%	0.94%	0.17%	Spk002	78.42%	12.41%	9.16%
Spk003	94.78%	4.54%	0.68%	Spk004	96.83%	2.05%	1.11%
Spk005	56.51%	26.11%	17.38%	Spk006	99.32%	0.51%	0.17%
Spk007	87.07%	7.36%	5.57%	Spk008	69.18%	17.72%	13.10%
Spk107	82.11%	9.59%	8.30%	Spk108	69.43%	18.15%	12.41%
Spk009	91.78%	5.31%	2.91%	Spk010	78.51%	13.10%	8.39%
Spk011	93.24%	5.15%	2.05%	Spk012	74.32%	13.96%	11.73%
Spk013	43.58%	30.48%	25.94%	Spk014	91.01%	5.14%	3.85%

Table 4: Labeling results per speaker

3. A total of 232 speakers were recorded in this subcorpus; with the balance of gender kept at 52% females and 48% males and ages were also balanced in the range of 10 to 18 years old. The final number of utterances was 13224 in 8h50m15s of data.

3.3. Human perceptual labeling at the lexical level

The major effect of the speakers disorders in their speech was at the lexical level, where they were producing a large number of inaccuracies and mispronunciations. To understand their influence in ASR systems, and to verify the results of pronunciation assessment tools, it was necessary to provide a ground truth for these mispronunciations.

Hence, a group of 14 labelers were chosen to obtain this manual labeling. For each isolated word utterance, 3 of these labelers were picked up to review that word and indicate for each phoneme if it had been correctly pronounced, substituted for another phoneme or completely deleted. A polling system decided the final label according to the majority of the votes from the 3 labelers, and a fourth labeler was requested for those cases where a decision could not be achieved among the 3 initial labelers. The labeling system was purely perceptual, this is, labelers, although people with experience in the fields of speech technologies or phonetics, were asked to annotate the signals only according to their own perception of the word. Furthermore, to avoid the effect of overtraining of the labelers perception to a given speaker, only a maximum of one session from a same speaker was handed to each labeler.

With this labeling system, the rates of correct, substituted and deleted phonemes were obtained for each speaker. The final rates for all the 3192 utterances from the disordered speakers showed that only 82.39% of the phonemes in the corpus were correctly pronounced, while 10.31% have been substituted and the remaining 7.30% deleted. Table 4 shows the rate of correct, substituted and deleted phonemes for each speaker, showing the large number of mispronunciations made by some speakers like *Spk013*, *Spk005*, *Spk008* or *Spk012*.

The pairwise interlabeler agreement rose up to 85.81%, which indicated the high consistency of the human annotators and validated their work, providing a useful ground truth for this task.

4. Experimental Results with the Corpus

This Section presents a series of results achieved with the corpus introduced in this paper, giving a quick glimpse to

all the possibilities that it opens for research in different tasks related to speech applications for the speech handi-capped.

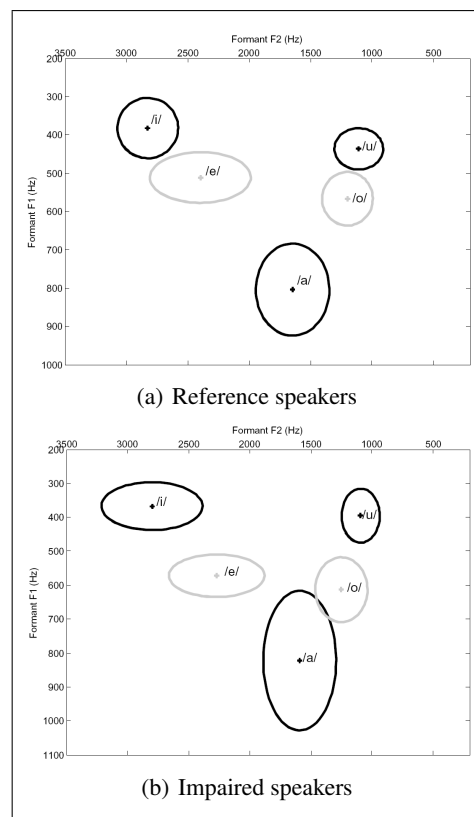


Figure 2: Estimated formant maps in the corpus

4.1. Acoustic and lexical analysis

A deep understanding of the special properties of the disordered speech is required previous to start developing any algorithm which deals with recognition or assessment of it. An analysis and measurement of the acoustic distortion of the speech production in the 5 Spanish vowels by the 14 impaired speakers was carried out (Saz et al., 2009c), showing the loss of formant separability between these speakers and their unimpaired age-matched peers.

Furthermore, the statistical processing of the lexical mispronunciations showed how the inaccuracies in these speakers were equivalent to that of preliterate young children in of 3 to 7 years old in the process of language acqui-

sition (Bosch-Galcerán, 2004). This could possibly indicate that the impaired speakers were suffering a big deal of language delay due to their multiple cognitive difficulties.

4.2. Speech recognition and adaptation

The present corpus is especially designed for the study of ASR systems with speaker adaptation. The presence of 4 sessions per speaker allows for the training of speaker specific models, both at the acoustic and the lexical level (Saz et al., 2009a).

The two most popular techniques for speaker adaptation: Maximum A Posteriori (MAP) (Gauvain and Lee, 1994) and Maximum Likelihood Linear Regression (MLLR) (Legetter and Woodland, 1995), were evaluated over the disordered speech corpus in a four series of leave-one-out experiments, achieving the results in Table 5. The baseline, with models trained from the unimpaired children subcorpus showed the pernicious influence of these disorders, as the results with the unimpaired speakers are 3% Word Error Rate (WER). The use of MAP or MLLR alone, can reduce this WER to 14-15% (a relative improvement of 45%), but the best results were achieved with an initial MLLR phase followed by MAP adaptation, achieving a 55% relative improvement.

Baseline	MAP	MLLR	MLLR+MAP
28.20%	15.48%	14.69%	12.53%

Table 5: Speech recognition results (WER) for the disordered speakers

4.3. Pronunciation assessment

Finally, the development of algorithms for pronunciation evaluation and assessment is also a major task which was expected to be covered by the corpus, due to the interests of the authors in Computer-Assisted Language Learning (CALL) tools. The results obtained have achieved promising results (Saz et al., 2009b; Yin et al., 2008; Yin et al., 2009), showing the possibilities of different techniques like score normalization or re-scoring based on lattices.

The different techniques evaluated have achieved results around a 15% of Equal Error Rate (ERR), which measures the point of the detection curve in which the number of false alarms equals the number of false rejects. The ground truth considered for the distinction of the correct pronounced phonemes and the mispronounced ones was the output of the human labeling shown in Section 3.. These results encourage the further research in this area for the development of CASLT tools which can really provide a useful feedback to the students and improve their capabilities in the oral language.

5. Conclusions and Future Work

Once a certain maturity of the corpus has been reached by the developers, it is strongly believed that the time has come to make the full corpus available for the research community. It is believed that the corpus can provide a valuable framework for the experimentation in disordered speech; belief which is sustained by the following key points:

- It contains sufficient data for significant experiments in different areas (more than 11 hours and a half of data).
- It characterizes a wide range of disorders in the range of ages of the impaired speakers, with a large group of unimpaired age-matched peers.
- It provides sufficient extra material to work like the full outcome of the annotation by the human experts.

Due to the special social interest that the research in HLTs for the handicapped has in itself, the “Alborada-I3A” corpus is freely available for every research group interested in the area. The only requirements to be fulfilled are two: To make use of the corpus purely for research purposes, as speech was freely donated by the speakers; and to keep the privacy of the speakers, as it contains speech from children under age and disabled individuals, whose privacy is protected under different laws. For that reason, no information is provided from the speakers, apart from their gender, age and diagnosis of their disorders.

Any group interested in the corpus can contact directly the authors via e-mail. The data for distribution contains all the speech signals from the impaired and unimpaired speakers, the outcome of the experts labeling as text files and different information files about the speakers. Its contribution to the LREC2010 Map is understood to help boosting the research in this area of interest.

6. Acknowledgements

This work was supported under national project TIN2008-06856-C05-04. The authors want to thank the staff from the CPEE “Alborada” (especially José Manuel Marcos and César Canalís), CEIP “Río Ebro” (Nuria Lozano), IES “Tiempos Modernos” (Pilar Burillo) and IES “Félix de Azara” (Joaquín Lostal) for their unique contribution to make this corpus and the whole work possible.

7. References

- L. Bosch-Galcerán. 2004. *Evaluación Fonológica del Habla Infantil*. Ed. Masson, Barcelona, Spain.
- S. Creer, S.-P. Cunningham, P.-D. Green, and K. Fatema. 2009. Personalizing synthetic voices for people with progressive speech disorders: Judging voice similarity. In *Proceedings of the 11th European Conference on Speech Communication and Technology (Eurospeech-Interspeech)*, pages 1427–1430, Brighton, UK, September.
- J.-R. Deller, M.-S. Liu, L.-J. Ferrier, and P. Robichaud. 1993. The whitaker database of dysarthric (cerebral palsy) speech. *Journal of the Acoustical Society of America*, 93(6):3516–3518.
- R. García-Gómez, R. López-Barquilla, J.-I. Puertas-Tera, J. Parera-Bermúdez, M.-C. Haton, J.-P. Haton, P. Alinat, S. Moreno, W. Hess, M.-A. Sánchez-Raya, E.-A. Martínez-Gual, J. L. Navas-Chabeli-Daza, C. Antoine, M.-M. Durel, G. Maurin, and S. Hohmann. 1999. Speech training for deaf and hearing impaired people:

- ISAEUS consortium. In *Proceedings of the 6th European Conference on Speech Communication and Technology (Eurospeech-Interspeech)*, pages 1067–1070, Budapest, Hungary, September.
- J.-L. Gauvain and C.-H. Lee. 1994. Maximum A Posteriori estimation for multivariate Gaussian mixture observations of Markov chains. *IEEE Transactions on Speech and Audio Processing*, 2(2):291–298.
- A. Hatzis, P. Green, J. Carmichael, S. Cunningham, R. Palmer, M. Parker, and P. O’Neill. 2003. An integrated toolkit deploying speech technology for computer based speech training with application to dysarthric speakers. In *Proceedings of the 8th European Conference on Speech Communication and Technology (Eurospeech-Interspeech)*, pages 2213–2216, Geneva, Switzerland, September.
- H. Kim, M. Hasegawa-Johnson, A. Perlman, J. Gunderson, T. Huang, K. Watkin, and S. France. 2008. Dysarthric speech database for universal access research. In *Proceedings of the 10th International Conference on Spoken Language Processing (ICSLP-Interspeech)*, pages 1741–1744, Brisbane, Australia, September.
- S. Lee, A. Potamianos, and S. Narayanan. 1999. Acoustics of children’s speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America*, 105(3):1455–1468.
- C.-J. Legetter and P.-C. Woodland. 1995. Maximum Likelihood Linear Regression for speaker adaptation of the parameters of continuous density Hidden Markov Models. *Computer Speech and Language*, 9:171–185.
- X. Menéndez-Pidal, J.-B. Polikoff, S.-M. Peters, J. Lorenzo, and H.-T. Bunnell. 1996. The Nemours database of dysarthric speech. In *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP-Interspeech)*, pages 1962–1965, Philadelphia (PA), USA, October.
- M. Monfort and A. Juárez-Sánchez. 1989. *Registro Fonológico Inducido (Tarjetas Gráficas)*. Ed. Cepe, Madrid, Spain.
- J.-L. Navarro-Mesa, P. Quintana-Morales, I. Pérez-Castellano, and J. Espinosa-Yáñez. 2005. Oral corpus of the project HACRO (help tool for the confidence of oral utterances). Technical report, Department of Signal and Communications, University of Las Palmas de Gran Canaria, May.
- M. Parker, S.-P. Cunningham, P. Enderby, M.-S. Hawley, and P.-D. Green. 2006. Automatic speech recognition and training for severely dysarthric users of assistive technology the stardust project. *Medical Engineering & Physics*, 20(2–3):149–156.
- O. Saz, E. Lleida, and A. Miguel. 2009a. Combination of acoustic and lexical speaker adaptation for disordered speech recognition. In *Proceedings of the 11th European Conference on Speech Communication and Technology (Eurospeech-Interspeech)*, pages 544–547, Brighton, United Kingdom, September.
- O. Saz, E. Lleida, and W.-R. Rodríguez. 2009b. Avoiding speaker variability in pronunciation verification of children disordered speech. In *Proceedings of the 2009 Workshop on Children, Computer and Interaction*, Cambridge (MA), USA, November.
- O. Saz, J. Simón, W.-R. Rodríguez, E. Lleida, and C. Vaquero. 2009c. Analysis of acoustic features in speakers with cognitive disorders and speech impairments. *EURASIP Journal on Advances in Signal Processing*, Special Issue on Analysis and Signal Processing of Oesophageal and Pathological Voices.
- O. Saz, S.-C. Yin, E. Lleida, R. Rose, W.-R. Rodríguez, and C. Vaquero. 2009d. Tools and technologies for computer-aided speech and language therapy. *Speech Communication*, 51(10):948–967.
- C. Vaquero, O. Saz, E. Lleida, and W.-R. Rodríguez. 2008. E-inclusion technologies for the speech handicapped. In *Proceedings of the 2008 International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4509–4512, Las Vegas (NV), USA, April.
- K. Vicsi, P. Roach, A. Oester, Z. Kacic, P. Barczikay, and I. Sinka. 1999. SPECO: A multimedia multilingual teaching and training system for speech handicapped children. In *Proceedings of the 6th European Conference on Speech Communication and Technology (Eurospeech-Interspeech)*, pages 859–862, Budapest, Hungary, September.
- S.-C. Yin, R. Rose, O. Saz, and E. Lleida. 2008. Verifying pronunciation accuracy from speakers with neuromuscular disorders. In *Proceedings of the 10th International Conference on Spoken Language Processing (ICSLP-Interspeech)*, pages 2218–2221, Brisbane, Australia, September.
- S.-C. Yin, R. Rose, O. Saz, and E. Lleida. 2009. A study of pronunciation verification in a speech therapy application. In *Proceedings of the 2009 International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4609–4612, Taipei, Taiwan, April.