A Novel Corpus of Children's Disordered Speech

Oscar Saz, William Rodríguez, Eduardo Lleida and Carlos Vaquero
Communications Technology Group (GTC)
Aragon Institute for Engineering Research (I3A)
University of Zaragoza, Zaragoza, Spain
{oskarsaz,wricardo,lleida,cvaquero}@unizar.es

ABSTRACT

This paper introduces the acquisition, evaluation and baseline Automatic Speech Recognition (ASR) experiments of a novel corpus containing speech from a set of impaired and unimpaired young speakers. A group of 14 speakers with different speech disorders have uttered several sessions over a 57-word vocabulary in Spanish to gather more than 3 hours of speech. In addition to this work, a parallel corpus of speech from unimpaired young speakers has been recorded with more than 6 hours of speech with the same vocabulary. The impaired speech corpus has been evaluated through a manual labeling to detect the mispronunciations made by the speakers, and the outcome of this work show that 17.31% of the phonemes have been either mispronounced or deleted in an isolated work task. A baseline evaluation of the performance of an state-of-the-art ASR system shows a 35.02% of Word Error Rate (WER) when using Speaker Independent models based on adult speech. This WER is reduced to 27.60% using models based on children speech and further reduced to 15.35% using speaker dependent models. Finally, experiments on connected speech show how ASR performance degrades on 4 impaired speakers on the transition from isolated words to connected speech due to the language impairments of the speakers and the coarticulation in connected speech.

1. INTRODUCTION

The use of Speech Technology-based systems for the improvement of the quality of life of the impaired ones is becoming a major line of research in the latest 20 years. Two main directions are taken to fulfill the objective of applying these Technologies: On one hand, the development of environment control systems based on oral interfaces for physically disabled individuals [6]; and, on the other hand, systems for providing computer-aided speech and language therapy to the speech handicapped [5]. One specific area is the application of all these systems to the population of physical or developmental disabled children. However, this research present gathers two major challenges for the use of Speech Technologies like Automatic Speech Recognition (ASR) or Pronunciation Verification (PV). First, it is well known that children's speech contains a set of specific parameters [8] that make them a group of users with special requirements in the use of Speech Technologies. Furthermore, speech disorders have another special issues when dealing with their speech in the field of Speech Technology-based systems [1].

Hence, the need of task oriented corpora for research arises. Regarding corpora that include disordered speech like dysarthric speech it is possible to find different examples in the literature [11, 17, 4, 15] for several languages (English, Dutch, Spanish...) and oriented to collect speech from adult speakers. This paper aims to expand the work in the collection of corpora in the area of impaired speech population. Specifically, this paper presents the acquisition and preliminary results with a database of young Spanish speakers suffering from different developmental and neuromuscular disorders that lead to several kinds of speech impairments like dysarthria.

The paper is organized as follows: Section 2 introduces the framework of the corpus collection and the goals in its acquisition. Sections 3 and 4 will present the speakers and sessions in the corpus. The two extra objectives in the work, collection of a reference children speech corpus and the manual labeling of the impaired speech corpus are presented in Sections 5 and 6 respectively. Finally, the baseline ASR results with the corpus are shown on Section 7, discussion to his results is given in Section 8 and conclusions to the work are obtained in Section 9.

2. FRAMEWORK AND REQUIREMENTS

The acquisition of the corpus introduced in this work is framed inside *Comunica*. *Comunica* intends to be a platform for the development and distribution of computer-aided speech therapy tools and research in Speech Technologies applied to handicapped people [16]. Being necessary to work with a corpus that reflect the special features in the speech of the target users of this assistive technology, the acquisition process started aiming to fulfill the following requirements:

- The corpus had to contain enough speech from young speakers and children, while keeping balance in terms of age and gender.
- Speakers had to suffer different kind and degrees of speech and language impairments caused by development and neuromuscular disorders.
- Their speech had to be collected in a realistic way that reflects all the features in the speakers' speech.
- The vocabulary used had to be short but containing the most prominent features of the Spanish language.
- Several sessions per speaker have to be collected to model intraspeaker variability.

Speakers and their diagnosis								
Speakers			Diagnosis					
			Disorder		Speech Disorder	Language disorder		
Speaker	Age	Gender	Developmental	Neuromuscular	Dysarthia	Semantic	Syntax	Other
*Spk*01	14yo	Female	Down's Syndrome	No	No	Yes	No	
*Spk*02	11yo	Male	Yes	No	No	Yes	Yes	
*Spk*03	21yo	Male	Yes	Yes	Yes	Yes	No	
*Spk*04	21yo	Female	Yes	No	No	Yes	No	
*Spk*05	18yo	Male	Down's Syndrome	Yes	Yes	Yes	Yes	Dysphemia
*Spk*06	17yo	Male	Yes	Ataxia	Yes	Yes	No	
*Spk*07	18yo	Male	Severe	Multiple	Yes	Yes	No	
*Spk*08	19yo	Male	Severe	Cerebral palsy	Yes	Yes	Yes	
*Spk*09	11yo	Female	Yes	Yes	Yes	No	No	
*Spk*10	15yo	Female	Yes	Yes	Yes	Yes	Yes	
*Spk*11	20yo	Female	Yes	Yes	Yes	Yes	No	
*Spk*12	18yo	Male	Severe	Yes	Yes	Yes	Yes	
*Spk*13	13yo	Female	Down's Syndrome	Yes	Yes	Yes	Yes	
*Spk*14	11yo	Female	Yes	Yes	Yes	No	No	

Table 1: Speakers in the impaired speech corpus

3. SPEAKERS IN THE CORPUS

The corpus presented in this work contains speech from 14 young speakers, 7 males and 7 females with ages ranging from 11 to 21 years old. Table 1 show the age and gender of the 14 speakers as well as a summary of their disorders and impairments. As it can be seen, all the speakers are affected by developmental impairments (sometimes as severe as Down's syndrome) and in some ways of neuromuscular disorders like cerebral palsy or ataxia. In terms of speech disorders, degrees of dysarthria are suffered by 11 of the speakers. Dysarthria is a speech impairment associated to neuromuscular disorders in which the lack of control of some of the phonatory and articulatory organs leads to the mispronunciation of one or several of the phonemes in the language. All of the speakers suffer from language disorders that affect the semantic and syntax abilities of their communication.

4. SESSIONS IN THE CORPUS

There are three main aspects in the design of the recording sessions: The vocabulary, the number of sessions and the technical aspect of the recordings.

4.1 Vocabulary

The recordings were oriented to make the sessions short and comfortable for the speakers (as the speaker population are children with disabilities) while trying to obtain the most complete set of sounds possible and a relatively large amount of data. The vocabulary chosen for the recordings is based in the 57-word set included in the Induced Phonological Register (RFI) [12], which is a well-known handbook for speech therapy in Spanish. RFI, while containing 57 words, contains examples of all the 23 phonemes and nearly every allophone of the 51 allophones described traditionally in the Spanish language [10]. The total amount of syllables in the 57 words is 129 (an average of 2,26 syllables per word, with 90 different syllables) and the number of phonemes is 292 (an average of 5,13 phonemes per word). The whole set of words and their phonetic transcriptions in SAMPA notation are shown on Table 2.

4.2 Number of sessions

The sessions design aimed to create short sessions as indicated previously. Hence, isolated-word recording sessions were the basis of the acquisition scenarios. Every speaker was told to utter 4 sessions of the 57 words in the RFI for a total of 228 utterances. Four extra sessions with 28 short simple sentences were also recorded by speakers *Spk*01, *Spk*04, *Spk*06 and *Spk*11. These simple sentences are meaningless sentences following a structure like this:

$$el/la\ Word1\ y\ el/la\ Word2$$

Where *el/la* is the article (*the* in English) for masculine and feminine words respectively and *y* is the copulative conjunction (*and* in English). *Word*1 and *Word*2 are two different words from the RFI. Every word appears only once in every one of the 4 simple-sentence sessions, and from the 4 times that every word is uttered by every speaker, two are in the initial position (as *Word*1) and two in the final position (as *Word*2). A final session containing 10 long complex sentences with complete meaning and containing three words from the RFI was recorded by Speakers *Spk*04, *Spk*06 and *Spk*11.

Every session was recorded in different days for a better modeling of intra-speaker variability. The total amount of speech data is 3,192 isolated-word utterances, (2 hours and 56 seconds of speech including silence) 459 short-sentence utterances (25 minutes and 31 seconds of speech including silence) and 30 long-sentence utterances (2 minutes and 9 seconds of speech including silence).

4.3 Technical aspect

Speech acquisition was carried out in the facilities of the School for Special Education (CPEE) "Alborada", located in Zaragoza (Spain) to make the speakers feel comfortable with the recording environment. *Vocaliza* [18] was the recording tool used in this corpus as this application includes a recording option and provides a friendly graphic environment for the speaker while reinforcing the correct pronunciation of the word or sentence via text and synthesized audio. From the technical aspect, a wireless close-talk microphone (model AKG C444L) was used in the recording with a commercial

Set of words and transcriptions in the corpus							
Word	SAMPA	Word	SAMPA	Word	SAMPA	Word	SAMPA
árbol	/aɾbol/	boca	/boka/	bruja	/bɾuxa/	cabra	/kabɾa/
campana	/kam̪pana/	caramelo	/kaɾamelo/	casa	/kas̪a/	clavo	/klabo/
cuchara	/kut͡ʃaɾa/	dedo	/dedo/	ducha	/dut͡ʃa/	escoba	/eskoba/
flan	/flan/	fresa	/fɾesa/	fuma	/fuma/	gafas	/gafas/
globo	/globo/	gorro	/goɾo/	grifo	/gɾifo/	indio	/indio/
jarra	/xaɾa/	jaula	/xaula/	lápiz	/lapiθ/	lavadora	/labadoɾa/
luna	/luna/	llave	/ʎabe/	mariposa	/maɾiposa/	moto	/moto/
niño	/niɲo/	ojo	/oxo/	pala	/pala/	palmera	/palmeɾa/
pan	/pan/	peine	/peine/	periódico	/peɾiodiko/	pez	/peθ/
piano	/piano/	pie	/pie/	piña	/piɲa/	pistola	/pistola/
plátano	/platano/	playa	/plaja̠/	preso	/pɾeso/	pueblo	/pueblo/
puerta	/puerta/	ratón	/raton̠/	semáforo	/semaforo/	silla	/siʎa/
sol	/sol/	tambor	/tambor/	taza	/taθa/	teléfono	/telefono/
toalla	/toaʎa/	toro	/toɾo/	tortuga	/toɾtuga/	tren	/tɾen/
zapato	/θapato/						

Table 2: Words in the Induced Phonological Register and their SAMPA transcription

Percentage of correct, mispronounced and deleted phonemes in the corpus							
Speaker	Correct	Mispronounced	Deleted	Speaker	Correct	Mispronounced	Deleted
*Spk*01	98.88%	0.94%	0.17%	*Spk*02	78.42%	12.41%	9.16%
*Spk*03	94.78%	4.54%	0.68%	*Spk*04	96.83%	2.05%	1.11%
*Spk*05	56.51%	26.11%	17.38%	*Spk*06	99.32%	0.51%	0.17%
*Spk*07	87.07%	7.36%	5.57%	*Spk*08	69.18%	17.72%	13.10%
*Spk*09	91.78%	5.31%	2.91%	*Spk*10	78.51%	13.10%	8.39%
*Spk*11	93.24%	5.15%	2.05%	*Spk*12	74.32%	13.96%	11.73%
*Spk*13	43.58%	30.48%	25.94%	*Spk*14	91.01%	5.14%	3.85%
Mean	82.39%	10.31%	7.30%				

Table 3: Results of the labeling process

use laptop with a internal sound card. This way, comfortability of the speaker was guaranteed as they were not directly attached to the computer while obtaining the best speech quality possible with a Signal-to-Noise Ratio (SNR) of 26.35 dB.

5. REFERENCE CORPUS

As a parallel process to the recordings of the impaired speech corpus, the recording of a reference corpus containing speech from unimpaired speakers in the same range of ages that the impaired speakers was made. This reference corpus was considered necessary to avoid the mismatch due to the difference of age between the impaired speakers and the age of the speakers contained in the usual adults' speech corpora used for training speaker independent models for ASR. Avoiding this mismatch is important since it could mask the mismatch due to the impairments of the speakers that gets better reflected when comparing speech from subjects in the same range of age. It also allows the research in Speech Technologies for unimpaired children in Spanish.

This unimpaired speech corpus should repeat the same acquisition scenario that the impaired speech corpus. Thus, the same vocabulary (RFI) and the same type of sessions (isolated words) were designed as the recording scenario. The number of speakers in this corpus is 168, 73 boys and 95 girls ranging in ages from 10 to 18 years old. Every speaker uttered a session of the 57 words in the RFI, which makes a total number of 9,576 isolated-word utterances in the corpus (6 hours, 17 minutes and 43 seconds of speech including silence) with an average SNR of 25.59 dB.

6. MANUAL LABELING OF MISPRONUNCIATIONS

Once the corpus was acquired, the evaluation of any robust ASR or pronunciation verification algorithm requires of a manual labeling in the corpus that validates the systems. Thus, a labeling process was started. In this process, every phoneme in the database was labeled by three different labelers as having been either deleted, mispronounced and therefore substituted with another phoneme, or correctly pronounced. In the end, the final label for the phoneme was chosen by consensus among the labelers. The average percentage of correct phonemes is 82.39%, while 10.31% of the phonemes are substituted and 7.30% are deleted. Full results for every speaker are shown on Table 3. Further results are given in Table 4 where it is shown the percentage of correct, mistaken and deleted phonemes for every of the 23 Spanish phonemes. Vowels like /a/, /o/ or /e/ are the ones are the easiest to pronounce by the speakers (over 90% of correctness), while consonants like /ɾ/, /ɾ/ or /θ/ represent the most challenging sounds for the speakers (less than 60% of correctness).

The strategy in the labeling resembles more a lexical labeling than a speech quality labeling. This was originally intended to create a more objective measure of the mispronunciations by the speakers in the corpus. Furthermore, it matches the speech therapy strategy used with this kind of severe disorders, where the goal is to provide the patient with the ability to distinguish phonemes correctly in the speech to lead to intelligibility in oral communication. Also,

Percentage of correct, mispronounced and deleted phonemes in the corpus									
Phoneme	Appearances	Correct	Mistaken	Deleted	Phoneme	Appearances	Correct	Mistaken	Deleted
/a/	3248	95.84%	2.37%	1.79%	/o/	2128	95.39%	3.52%	1.08%
/p/	1064	87.41%	7.80%	4.79%	/e/	1008	90.08%	5.75%	4.17%
/ɾ/	1008	50.80%	21.63%	27.58%	/l/	952	61.45%	17.96%	20.59%
/i/	752	83.93%	7.91%	8.16%	/t/	784	87.76%	8.16%	4.08%
/n/	672	84.08%	7.89%	8.04%	/b/	616	77.60%	17.21%	5.19%
/s/	560	79.46%	10.89%	9.64%	/u/	504	84.33%	10.71%	4.96%
/k/	504	76.79%	15.67%	7.54%	/m/	448	69.42%	12.50%	18.08%
/f/	392	75.26%	19.64%	5.10%	/d/	336	66.07%	22.62%	11.31%
/g/	280	61.43%	26.07%	12.50%	/j̟/	224	82.59%	15.18%	2.23%
/θ/	224	58.48%	19.20%	22.32%	/x/	224	91.52%	6.70%	1.79%
/r/	168	35.12%	61.31%	3.57%	/ɲ/	112	86.61%	7.14%	6.25%
/t͡ʃ/	112	63.39%	35.71%	0.89%					

Table 4: Results of the labeling process

Isolated Word ASR Results					
Speakr	Adult	Child	Speakr	Adult	Child
*Spk*01	2.19%	3.95%	*Spk*02	36.84%	16.23%
*Spk*03	27.19%	7.46%	*Spk*04	18.42%	4.39%
*Spk*05	62.28%	54.82%	*Spk*06	9.65%	5.26%
*Spk*07	28.95%	25.00%	*Spk*08	47.81%	45.61%
*Spk*09	32.89%	23.25%	*Spk*10	39.47%	32.46%
*Spk*11	13.60%	9.21%	*Spk*12	69.30%	62.28%
*Spk*13	77.63%	77.63%	*Spk*14	24.12%	18.86%
Mean	33.24%	26.38%			

Table 5: Speaker independent ASR results for adults and children speech models

Isolated Word ASR Results					
Speakr	Adult	Child	Speakr	Adult	Child
*Spk*01	2.19%	2.19%	*Spk*02	12.72%	10.53%
*Spk*03	2.19%	1.75%	*Spk*04	2.63%	2.19%
*Spk*05	44.30%	46.05%	*Spk*06	2.63%	1.75%
*Spk*07	5.70%	6.14%	*Spk*08	31.14%	25.44%
*Spk*09	11.84%	10.96%	*Spk*10	11.40%	12.28%
*Spk*11	3.95%	3.07%	*Spk*12	16.67%	17.98%
*Spk*13	63.60%	68.42%	*Spk*14	6.58%	6.14%
Mean	14.07%	14.13%			

Table 6: Speaker dependent ASR results for adults and children speech models

with this type of labeling, labels are more consistent as the pair-wise inter-labeler agreement rate is 85.81% which raises to 89.65% when considering only a binary decision: Correct versus Incorrect (deletions plus substitutions). This consistent labeling avoids the problems of a subjective speech quality measurement that would have required of very experienced labelers and would have suffered of subjective differences between the evaluation given by several labelers.

7. ASR BASELINE EXPERIMENTS

Finally, an evaluation of the performance of ASR over the corpus was made. Getting to know the influence of the speech impairments in the performance of ASR is an important issue to face the development of new Speech-Technology based systems. The experimental study is carried out over the impaired speech corpus acquired in this work. The results with an adult speaker independent model are obtained firstly; subsequently, the non-impaired speech part of the corpus is used to train a children speaker independent model and finally speaker dependent models are obtained for every one of the 14 impaired speakers to achieve the final results. All the acoustic models are based on a set of 746 context-dependent units plus a begin-end silence model ; every model being a 1-state Hidden Markov Model (HMM) whose distribution is a Gaussian Mixture Model (GMM) made up of 16 gaussians. Features for the ASR system are a 39-feature vector with the features extracted every 10 milliseconds using a 25-millisecond Hamming window; 12 first mel-frequency-cepstrum coefficients (MFCC) and the log-energy are then obtained and are used for the system plus the first and second derivatives of them; log-energy plus its first and second derivatives are also calculated to create a 39-feature vector.

7.1 Speaker independent results

Results for the speaker independent set of experiments are obtained for every one of the 14 impaired speakers. Two speaker independent models are trained: First one is obtained using the unimpaired adults' speech corpora Albayzin [14], SpeechDat-Car [13] and Domolab [7] via a Maximum Likelihood (ML) [2] approach. The second one is a task and domain adaptation carried out by creating an unimpaired children speaker independent model using the 9,576 isolated words in the unimpaired speakers corpus. The adaptation is carried out via Maximum A Posteriori (MAP) algorithm [3] because having such a large amount of data for adaptation provides a very good convergence of the algorithm, so no other algorithms like MLLR [9] are needed.

Results for both adult and children speech models are shown on Table 5. The average Word Error Rate (WER) for all the speakers using the adult speech model reaches a 33.24%, much higher than the results obtained using the same ASR system on the unimpaired children's corpus, whose average WER is 3.30%. These results show the big influence of the speakers' disorders, specially for the 3 most impaired speakers whose WER is above 60% (*Spk*05, *Spk*12 and *Spk*13). The average WER with the children speech model is 26.38% (which is a reduction of 22.92% in the WER) and reflects the mismatch due to the speakers' disorders when the model is adapted to the task (RFI words) and the domain (children's speech). The reference WER in this task is 2.18% doing an ASR experiment in which every half of the unimpaired children corpus is used to train a model tested on the other half and then averaging both results.

Connected Speech ASR Results					
Speakr	Adult	Child	Speakr	Adult	Child
*Spk*01	7.52%	7.63%	*Spk*04	23.01%	13.50%
*Spk*06	7.08%	4.42%	*Spk*11	15.93%	8.30%
Mean	13.39%	12.94%			

Table 7: Speaker independent ASR results for adults and children speech models

Connected Speech ASR Results					
Speakr	Adult	Child	Speakr	Adult	Child
*Spk*01	11.95%	7.08%	*Spk*04	19.47%	14.60%
*Spk*06	6.19%	3.87%	*Spk*11	14.16%	9.29%
Mean	8.46%	8.71%			

Table 8: Speaker dependent ASR results for adults and children speech models

7.2 Speaker dependent results

Speaker dependent ASR experiments were carried out within a leave-one-out strategy in which three of the isolated-word sessions of every speaker were used for adaptation and the remaining session is used for evaluation. The definitive WER for every speaker is obtained as the average WER for the 4 evaluated sessions and a MAP approach is used for the adaptation. Results for the speaker dependent set of experiments in Table 6 show that the average WER decreases to 14.07% and 14.13% (64.04% and 63.83% of improvement over the initial results) when using the adapted models with different seeds (adults' speaker independent and children's speaker independent) respectively.

7.3 Connected speech ASR results

Finally, the same experiments are run over the 459 short simple sentences recorded from 4 speakers. The speaker independent results with adult and children speech models are shown on Table 7 and the speaker dependent results with adult and children models as seed for adaptation are shown on Table 8. The recognition results are only related to the correct recognition of the RFI words and not to the connective words that are into every sentence. This way, the experiments resemble the recognition of two different isolated words in the sentence.

Average results show a 13.39% WER for speaker independent adult speech models and 12.94% for speaker independent children speech model (3.36% of reduction). While for speaker dependent models, 8.46% of WER is the result with an adult seed and 8.71% with children speech as seed (36.82% and 32.69% of WER reduction compared to speaker independent models).

8. DISCUSSION

The first matter of discussion in the ASR results shown in Section 7 is the relation between the results for the four main experiments carried out. Figure 1 shows graphically the steps in the WER reduction that the task and domain adaptation and furthermore the speaker adaptation produce. It is to notice that speaker adaptation produces a much bigger reduction in WER than task domain adaptation.

Another subject for discussion is the differences between ASR results of isolated words and connected speech. As

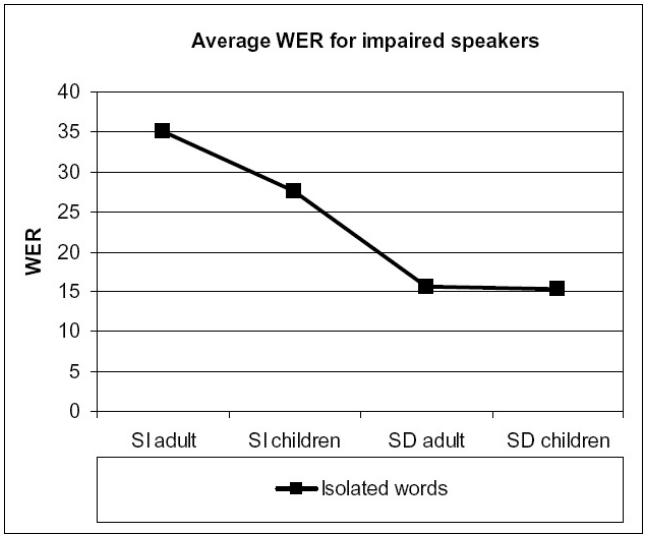

Figure 1: Average WER results

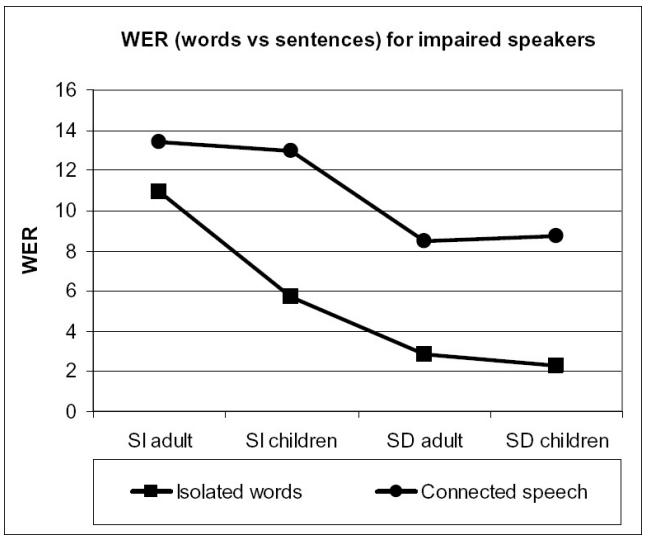

Figure 2: Isolated word vs connected speech ASR results

it can be seen on Figure 2, the results of connected speech are always worse that the average for the same 4 speakers in isolated words (*Spk*01, *Spk*04, *Spk*06 and *Spk*11). Considering that the recognition task is designed to be comparable to isolated words (the influence of the connecting words is neglected as they are forced in the recognition); it can be easily argued that this difference appears due to the language disorders of the speakers that adds to their speech disorders, making worse their pronunciation. Coarticulation effects and its difficulty in speech production for the speakers also relates to this effect.

9. CONCLUSIONS

As a result of this work, a novel corpus with unimpaired children's speech has been acquired. This corpus reflects the several articulation disorders of 14 young speakers affected by dysarthria. As complimentary work, a parallel reference corpus with unimpaired children's speech has been recorded for task domain adaptation on ASR-based systems; and a lexical labeling of the impaired corpus has been car-

ried out, where 17.62% of the phonemes in the corpus have been marked as either deleted or substituted. The performance of ASR systems has been checked over this corpus, and a heavy increase in the WER has been noticed between the unimpaired and the impaired speech. The WER can be decreased by the use of task domain adaptation with the unimpaired speech and a further decrease can be obtained applying speaker adaptation over the impaired speech signals. However, the final WER of 15.35% indicates that further research is required in acoustic and lexical modeling to make plausible the use of ASR systems with disordered speech young speakers. Finally, a heavy decrease in ASR performance (specially with speaker dependent models) has been detected in connected speech referred to isolated words due to coarticulation effects and the language impairments of the speakers.

10. ACKNOWLEDGMENTS

This work has been supported by the national project TIN-2005-08660-C04-01 from MEC of the Spanish government.

Authors want to thank Pedro Pegero, José Manuel Marcos and César Canalís from the CPEE "Alborada" in Zaragoza (Spain) for making possible this work. Also, special thanks to the people of the CEIP "Río Ebro" and the IES "Tiempos Modernos" in Zaragoza (Spain) for their collaboration.

11. REFERENCES

[1] J.-R. Deller, D. Hsu, and L.-J. Ferrier. On the use of Hidden Markov Modelling for recognition of dysarthric speech. *Computer Methods and Programs in Biomedicine*, 35:125–139, 1991.

[2] A.-P. Dempster, N.-M. Laird, and D.-B. Rubin. Maximum Likelihood for incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39(1):1–21, 1977.

[3] J.-L. Gauvain and C.-H. Lee. Maximum A Posteriori estimation for multivariate Gaussian mixture observations of Markov chains. *IEEE Transactions on Speech and Audio Processing*, 2(2):291–298, 1994.

[4] P. Green, J. Carmichael, A. Hatzis, P. Enderby, M. Hawley, and M.Parker. Automatic Speech Recognition with sparse training data for dysarthric speakers. In *Proceedings of the 8th European Conference on Speech Communication and Technology (Eurospeech-Interspeech)*, Geneva, Switzerland, September 2003.

[5] A. Hatzis, P. Green, J. Carmichael, S. Cunningham, R. Palmer, M. Parker, and P. O'Neill. 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)*, Geneva, Switzerland, September 2003.

[6] M. Hawley, P. Enderby, P. Green, S. Brownsell, A. Hatzis, M. Parker, J. Carmichael, S. Cunningham, P. O'Neill, and R. Palmer. STARDUST Ű Speech Training And Recognition for Dysarthric Users of aSsistive Technology. In *Proceedings of the 7th Conference of the Association for the Advancement of Assistive Technology in Europe, AAATE*, Dublin, Ireland, August 2003.

[7] R. Justo, O. Saz, V. Guijarrubia, A. Miguel, M.-I. Torres, and E. Lleida. Improving dialogue systems in a home automation environment. In *Proceedings of the First International Conference on Ambient Media and Systems (Ambi-Sys 2008)*, Québec City, Canada, February 2008.

[8] S. Lee, A. Potamianos, and S. Narayanan. Acoustic of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustic Society of America*, 105(3):1455–1468, March 1999.

[9] C.-J. Legetter and P.-C. Woodland. Maximum likelihood linear regression for speaker adaptation of the parameters of continous density hidden markov models. *Computer Speech and Language*, 9:171–185, 1995.

[10] E. Martínez-Celdrán. *Fonología General y Española: Fonología Funcional*. Ed. Teide, Barcelona, Spain, 1989.

[11] X. Menéndez-Pidal, J.-B. Polikoff, S.-M. Peters, J. Lorenzo, and H.-T. Bunnell. The Nemours database of dysarthric speech. In *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP-Interspeech)*, Philadelphia (PA), USA, October 1996.

[12] M. Monfort and A. Juárez-Sánchez. *Registro Fonológico Inducido (Tarjetas Gráficas)*. Ed. Cepe, Madrid, Spain, 1989.

[13] A. Moreno, B. Lindberg, C. Draxler, G. Richard, K. Choukri, S. Euler, and J. Allen. Speech dat car. a large speech database for automotive environments. In *Proceedings of the II Language Resources European Conference*, Athens, Greece, June 2000.

[14] A. Moreno, D. Poch, A. Bonafonte, E. Lleida, J. Llisterri, J.-B. M. no, and C. Nadeu. Albayzin speech database: Design of the phonetic corpus. In *Proceedings of the 3th European Conference on Speech Communication and Technology (Eurospeech-Interspeech)*, Berlin, Germany, September 1993.

[15] J.-L. Navarro-Mesa, P. Quintana-Morales, I. Pérez-Castellano, and J. E.-Y. nez. 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 2005.

[16] W. Rodríguez, O. Saz, E. Lleida, C. Vaquero, and A. Escartín. COMUNICA - Tools for speech and language therapy. In *Proceesing of the Workshop on Child, Computer and Interaction*, Chania, Greece, October 2008.

[17] E. Sanders, M. Ruiter, L.Beijer, and H. Strik. Automatic Recognition of Dutch dysarthric speech: A pilot study. In *Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP-Interspeech)*, Denver (CO), USA, September 2002.

[18] C. Vaquero, O. Saz, E. Lleida, and W.-R. Rodríguez. E-inclusion technologies for the speech handicapped. In *Proceedings of the 2008 International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas (NV), USA, April 2008.