Acta Universitaria
DIRECCIÓN DE APOYO A LA INVESTIGACIÓN Y AL POSGRADO

Communication Interface for Mexican Spanish Dysarthric Speakers.

Gladys Bonilla-Enriquez* and Santiago-Omar Caballero-Morales*

ABSTRACT

Dysarthria is a motor speech disorder due to weakness or poor coordination of the speech muscles. This condition can be caused by a stroke, cerebral palsy, or by a traumatic brain injury. For Mexican people with this condition there are few, if any, assistive technologies to improve their social interaction skills. In this paper we present our advances towards the development of a communication interface for dysarthric speakers whose native language is Mexican Spanish. We propose a methodology that relies on (1) special design of a training normal-speech corpus with limited resources, (2) standard speaker adaptation, and (3) control of language model perplexity, to achieve high Automatic Speech Recognition (ASR) accuracy. The interface allows the user and therapist to perform tasks such as dynamic speaker adaptation, vocabulary adaptation, and text-to-speech (TTS) synthesis. Live tests were performed with a mildly dysarthric speaker, achieving accuracies of 93%-95% for spontaneous speech.

RESUMEN

La disartria es una discapacidad motora del habla caracterizada por debilidad o poca coordinación de los músculos del habla. Esta condición puede ser causada por un infarto, parálisis cerebral, o por una lesión severa en el cerebro. Para mexicanos con esta condición hay muy pocas, si es que hay alguna, tecnologías de asistencia para mejorar sus habilidades sociales de interacción. En este artículo presentamos nuestros avances hacia el desarrollo de una interfaz de comunicación para hablantes con disartria cuya lengua materna sea el español mexicano. La metodología propuesta depende de (1) diseño especial de un corpus de entrenamiento con voz normal y recursos limitados, (2) adaptación de usuario estándar, y (3) control de la perplejidad del modelo de lenguaje para lograr alta precisión en el Reconocimiento Automático del Habla (RAH). La interfaz permite al usuario y terapéuta el realizar actividades como adaptación dinámica de usuario, adaptación de vocabulario, y síntesis de texto a voz. Pruebas en vivo fueron realizadas con un usuario con disartria leve, logrando precisiones de 93%-95% para habla espontánea.

Recibido: 10 de Enero del 2010
Aceptado: 14 de Febrero del 2012

INTRODUCTION

There is research in Mexico focused on developing Automatic Speech Recognition (ASR) technologies and Natural Language Processing (NLP). Within this context, we look towards developing a technological tool to assist Mexican people with disabilities. According to INEGI, in the year 2010, approximately 5.1% of the total Mexican population had some kind of disability, and within this percentage, 50.3% had a motor disability and 8.3% had a speech or communication disability.

Dysarthria is a motor speech disability that is often associated with irregular phonation and amplitude, in-coordination and restricted movement of speech articulators. This can be caused by a stroke or injury that affect the central nervous system, or by neuronal degenerative diseases like Multiple Sclerosis or Parkinson's disease [1]. Hence, dysarthria includes motor dysfunction of respiration, phonation, resonance, and articulation [2]. The type and severity of dysarthria depend on the part of the nervous system which is affected speakers.

Palabras clave:
Habla disártica; tecnología asistiva; reconocimiento automático del habla.
Keywords:
Dysarthric speech; assistive technology; automatic speech recognition.

*División de Posgrado, Universidad Tecnologica de la Mixteca, Carretera a Acatlima Km 2.5, C.P. 69000, Huajuapan de Leon, Oaxaca, Mexico. E-mail: gbonillae@live.com.mx, scaballero@mixteco.utm.mx

Acta Universitaria
DIRECCIÓN DE APOYO A LA INVESTIGACIÓN Y AL POSGRADO

Hence, there is a wide variety of abnormalities among dysarthric speakers. As a result of these dysfunctions, dysarthric speech is characterized by the following specific symptoms in the pronunciation of phonemes in the Spanish language [2]:

- Substitution: a phoneme is replaced by another. Most frequent: /*r*/ by /*d*/, /*s*/ by /*z*/, and /*k*/ by /*t*/.
- Deletion: a phoneme is omitted (e.g., "iño" is uttered instead of "niño", or "loj" is uttered instead of "reloj").
- Insertion: a phoneme that does not match the spoken phonemes is inserted to support the pronunciation of a phoneme which is difficult to utter (e.g., "Enerique" is uttered instead of "Enrique").
- Distortion: a sound that doesn't match a phoneme is uttered as a posible pronunciation for a phoneme with articulation difficulties.

There has been research for the development of technological tools to support people with this disability, specially in the field of ASR research. The use of commercial systems as Dragon Naturally Speaking, Microsoft Dictation, Infovox, etc. [1, 3] have shown varying levels of recognition in the range of 50% to 95% for users with different levels of dysarthria, obtaining the best performance for small vocabularies (10 - 78 words). Research projects have also been developed. In [4] the use of ANNs was explored, which performed better than the commercial system IntroVoice. Significant performances were also obtained in [5] with Hidden Markov Models (HMMs) for Dutch speaking users. In [6] accuracy rates of HMM-based ASR of 86.9% were obtained for British speakers with severe dysarthria and a vocabulary of 7-10 words.

However, work is limited for Mexican Spanish speakers. Significant research was found in [7]. Here, a phoneme processing system for rehabilitation of people with disordered speech using machine learning techniques was developed. This system consisted of modules that allowed the therapist to manage registration and therapy activities for patients (e.g., pronunciation exercises, audiovisual information).

However, to build a robust ASR system, usually a large training speech corpus is needed. Speech corpora are expensive and require long time to produce as each speech sample must be labelled at the phonetic and orthographic levels. Mexican Spanish corpora is limited, being the most significant the DIME corpus [8]. However this resource was not available as it is currently in licensing procedures. Also, large corpora from dysarthric speakers require more time and effort given their disability.

Hence, we explored on the use of a base ASR system - trained with the speech of a normal speaker - adapted to the speech patterns of a dysarthric speaker. Although this approach is general and well known for recognition of small sets of words (e.g., digits) [9], this hasn't been proved with disordered speech. Also, we considered that by designing a special text stimuli for production of training speech samples (and adaptation), a robust base ASR system can be accomplished. Also, we explored on the effect of continuous adaptation, Language Model (LM) perplexity, and control of LM restrictions, to improve ASR performance and achieve accuracies comparable to those obtained by commercial systems for a larger vocabulary. The realization of these exploratory studies led to the development of an ASR interface that achieved ASR accuracies of 93%-95% for a dysarthric speaker.

In this paper we present our findings and the development details of the interface as follows: in Section *Training Speech Corpus* the development of the training speech corpus for the base ASR system is presented; in Section *Interface Development* the construction details of each module of the interface (see figure 1) and how the ASR is managed to accomplish the respective tasks are presented; in Section *Performance Tests* information about the interface's tests is presented; finally, in Section *Conclusion* we discuss on our findings and the future work.

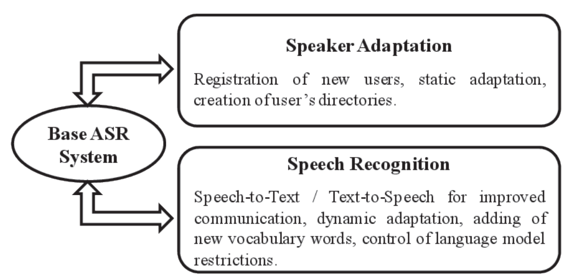

Figure 1 . Main modules of the proposed interface for dysarthric speakers.

TRAINING SPEECH CORPUS

To develop the ASR engine with limited resources it was assumed that robustness could be accomplished if: (1) there were enough speech samples (of all phonemes in the language) in the training speech corpus for acoustic modelling (even if only a single speaker were used as speech source); (2) the vocabulary were not large (< 1000 words); (3) the effect of statistical a-priori information (such as that of the LM) were controlled during ASR; and (4) dynamic speaker adaptation were performed while using the system.

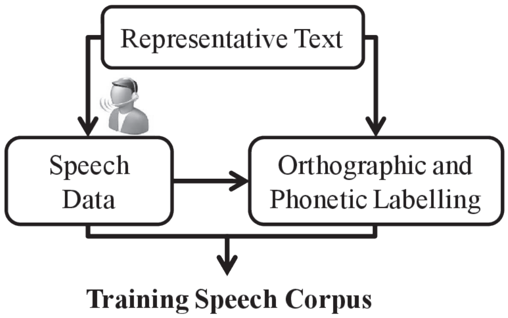

Figure 2 . Steps to obtain the speech corpus for supervised training of the base ASR system.

In figure 2 the steps followed to obtain the training speech corpus are shown. The representative text was obtained from the following sources:

- 49 words used for assessment of dysarthria in Mexican people. These words were provided by a speech therapist from the local National System for Integral Family Development (DIF) center.

- A fragment of the story "Fiesta en la Montaña" which was phonetically balanced and consisted of 102 words.

- 16 specially designed phonetically-balanced sentences. For new users, this text was the stimuli to obtain speech adaptation data.

Thus, the representative text consisted of 205 different words. The phoneme sequences that define each word were obtained with the tool *TranscribeMex* [8] which was developed to phonetically label the DIME corpus. Labelling of speech data at the phonetic level is important for acoustic modelling of phonemes.

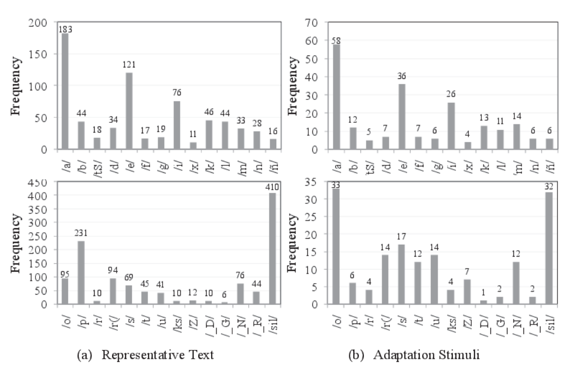

Figure 3 . Frequency of Mexican phonemes in the different stimuli data.

A total of 28 phonemes were identified and their frequency in the representative text is shown in figure 3(a). Based on [10], where for command recognition of speakers with severe dysarthria a minimum of six samples of a word was enough to get accuracies up to 100%, our representative text was considered to be well balanced to provide enough speech samples of a phoneme.

To obtain the speech data for the corpus the text was read 5 times by two male speakers: (1) a reference speaker with normal speech, and (2) a speaker with a low-medium diagnosed level of dysarthria (see section *Experiments with Dysarthric Speech*, table 1). This was performed to test two development methodologies used for ASR of dysarthric speech: (1) use of a Speaker - Dependent (SD) system (trained with dysarthric speech data from the user that will use the system) [4, 11, 6]; and (2) use a Speaker - Independent (SI) system (trained with speech data from a normal speaker) adapted to the speech of the dysarthric speaker who will use the system [3, 1]. The performance of these systems is discussed in section *Performance Tests*. The speech was recorded with a Sony lcd-bx800 recorder with a sampling frequency of 8 kHz monoaural in WAV format. This data was then labelled manually at the word (orthographic) and phonetic level with the tool *WaveSurfer*. With the realisation of this step the training corpus was finished and ready for the development of the base ASR system (see figure 4).

INTERFACE DEVELOPMENT

Base ASR System

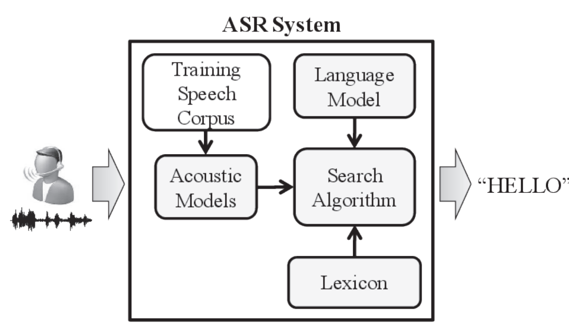

Figure 4 . Functional elements of an ASR system.

Acoustic Models

The acoustic models are the pattern recognition core of the ASR system, and are initialized and re-estimated with the data of the training corpus (supervised training of the ASR system). The technique used for acoustic modelling was HMMs [9], and the implementation tool was *HTK Toolkit* [9]. A HMM was built for each of the 28 phonemes of figure 3. These HMMs had standard three-state left-to-right architecture. A continuous probability distribution, that models the observation probability of a given acoustic signal (e.g., a phoneme), is associated to each state in a HMM. Thus, each phoneme has different probability distribution [9]. The observation probabilities are modelled as a mixture of gaussian distributions. While is common practice to use three gaussian components [6], performance of the ASR system is affected by the number of such components [9]. For ASR of dysarthric speech (and limited training data), this is considered a main factor. Thus, the **number of gaussian components** was considered as the **first variable** of the interface.

For supervised training, the corpus was coded into Mel Frequency Cepstral Coefficients (MFCCs). The front-end used 12 MFCCs plus energy, delta, and acceleration coefficients [9]. Once the user sets a number of gaussian components, the supervised training with the coded speech is performed by the interface.

Language Model and Lexicon

The Language Model (LM) represents a set of probabilities that restricts the recognised sequence of words from the ASR system to valid sequences, guiding the search algorithm to find the most likely sequence of words that best represent an input speech signal. Commonly, N-grams are used for the LM, and for this work, bigrams (N=2) were used for continuous speech recognition [9].

Two metrics are commonly used to measure the performance of a LM: (1) Word Error Rate, WER (1-%ASR Accuracy), and (2) perplexity. In some cases, low WER correlates with low perplexity [5]. For dysarthric ASR, low perplexity is recommended to deal with the effect of slow articulation of speech [5]. Perplexity is not ASR dependent, and thus, can be estimated faster than WER [12]. Perplexity increases as the vocabulary grows in size, and the use of a N-gram LMs reduces perplexity for large vocabularies as it restricts the possible sequences of words to most likely sequences. However to accomplish this, the test vocabulary must be known in advance by the ASR system [12].

To deal with this issue, we considered to build the LM on-line while using the interface, thus constantly updating the LM to allow advanced knowledge of the test vocabulary. Hence, the **vocabulary** and the **LM** were considered as a **second variable**. In addition, a **third variable** was considered, the LM's **scale grammar factor** (SGF). As this factor increases, the recogniser relies more on the LM instead of the acoustic signal to predict what the speaker said (e.g., the LM restrictions have more importance)[9]. Hence, the SGF can be used to reduce the perplexity of the LM during speech recognition. Thus, to accomplish control of the LM's perplexity, the following functions in the module were implemented: (1) manipulation of SGF; (2) cumulative estimation of bigrams (LM) considering each word, or sequence of words, added to the system.

Finally, the Lexicon specifies the sequences of phonemes that form each word in the application's vocabulary. Because the ASR is built at the phoneme level, at recognition time, the speech is decoded into sequences of phonemes which are restricted to form valid words by the Lexicon (which then are restricted by the LM to form valid word sequences). Each time a word is added to the LM the Lexicon is updated by managing the TranscribeMex tool.

Search Algorithm

The Viterbi algorithm is widely used for speech recognition [9]. This task consists in finding (searching) the sequence of words that best match the speech signal. While Viterbi decoding is implemented by the HTK library, it requires the following elements for decoding: (1) MFCC-coded speech to be recognised, (2) acoustic models, (3) LM, (4) lexicon, (5) SGF, and (6) list of phonemes. The interface manages the construction of each one of these elements and the execution of the Viterbi algorithm to recognise speech (details in section *Speech Recognition Module).*

Speaker Adaptation Module

Commercial ASR systems are trained with hundreds or thousands of speech samples from different speakers. When a new user wants to use such system, it is common to ask the user to read some words or texts (stimuli) to provide speech samples that are then used by the system to adapt its acoustic models to the patterns of the new user's voice. For this work, Maximum Likelihood Linear Regression (MLLR) [9] was the adaptation technique. MLLR is based on the assumption that a set of linear transforms can be applied to the parameters of the gaussian components of the ASR system's HMMs to reduce the mismatch between them and the adaptation data. A regression class tree with 32 terminal nodes was used for the dynamic implementation of MLLR [9]. 16 phonetically - balanced sentences were designed to be stimuli for this task (see figure 5). The frequency of phonemes of these sentences is shown in figure 3(b). The distributions of figure 3 correlate with each other with a coefficient of 0.62. In figure 5 we present the interface for the adaptation module. The programming language was *Matlab 2008* with the *GUIDE* toolkit.

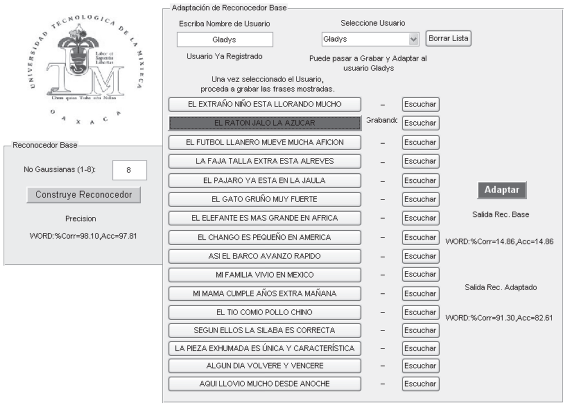

Figure 5 . Speaker Adaptation Module.

In the panel *Reconocedor Base*, the user can build the base ASR with different number of gaussian components. First the user must set the number of gaussian components in *No. Gaussianas (1-8)* and then press the button *Construye Reconocedor*. With this, the ASR system is built in just one step, and the user can evaluate its performance with different gaussian components. For this case, a maximum of 8 components was considered.

When the base ASR system is built, the user can access the panel *Adaptación de Reconocedor Base*, so it can be adapted to his/her voice. If the user is "new" then he/she must write his/her name in the field *Escriba Nombre de Usuario*. When the user finishes this task, the interface automatically saves the name and updates the list of registered users which is shown in the pop-up menu *Seleccione Usuario*. When the user selects his/her name from this list/menu, the user's directories are created (or loaded) for the adaptation (or re-adaptation) of the ASR system.

To start adaptation the user must record all 16 sentences shown in the push buttons. These can be recorded in any order. When the user presses a sentence's button, this turns "red", which means that the interface is recording the user's speech. When the user finishes reading the stimuli, then he/she must press again the sentence's button, which then turns "white". Specially for dysarthric speakers, the reading speed can be slow, so enabling the user to record speech with variable length was a priority for the design of the interface. There is a button next to each sentence's button labelled as *Escuchar*. This is to allow the user to listen his/her speech sample in order to verify if it was recorded correctly (and re-record it if needed). At the end of the recording tasks, the user just needs to press the button *Adaptar*. By doing this, the interface automatically creates (or loads) the personalised MLLR directories to create (or re-estimate) the adaptation transformations for that user, codecs the speech data into MFCCs, and performs MLLR. The accuracy results **(%Corr, Acc)** (see section *Performance Tests*) of the base ASR system on the adaptation data, before (in *Salida Rec. Base*) and after MLLR adaptation (in *Salida Rec. Adaptado*), are shown for comparison purposes.

Note that this kind of adaptation is usually performed once, before the new speaker uses the system (e.g., static adaptation). In commercial systems, if the speaker wishes to improve adaptation, he/she needs to read other stimuli texts. For our system we incorporated this task within the use of the ASR system, so adaptation can be made while performing speech recognition (e.g., dynamic adaptation). See section *Speech Recognition Module*.

Speech Recognition Module

Figure 6 . Speech Recognition Module.

The ASR interface for communication of dysarthric speakers is shown in figure 6. The user initially must select his/her name in the *Seleccione Usuario* pop-up menu. When selecting the user's name his/her adapted acoustic models are automatically loaded. There is also the button *Crea Modelo de Lenguaje* which builds the ASR's LM considering the vocabulary words/sentences displayed in *Frases de Vocabulario*. This is an informative list about the vocabulary stored in the system and available to be recognised. This interface allows the user to add more vocabulary to this list and thus reduce perplexity (see section *Language Model and Lexicon)*.

The new vocabulary must be typed in *Añadir Nuevas Frases o Palabras* in UPPER case format. By pressing the *Crea Modelo de Lenguaje* button, the interface updates the ASR's lexicon (by managing TranscribeMex) and the LM. Another parameter that can be set is the SGF (see section *Language Model and Lex-*

icon) to increase the influence of the LM in the recognition process. The SGF's value can be set in *Valor de Ajuste (1-30)*. The range for the SGF was set to the range 1-30 as it was observed that, for dysarthric speakers, maximum ASR accuracy is achieved with values over 20 [13]. For normal speech, usually a value of 5 is used [9]. In the interface this parameter can be changed at any moment without the need to re-start the system. Thus, ASR performance can be adjusted in real time.

To perform ASR for communication, the user must press the button *Ejecuta Reconocedor*. This button turns to "red" when starts recording and turns back to "white" when pressed again to finish the recording process. Internally, when recording finishes, the interface performs parametrization of the recorded speech, manages the HTK library to perform Viterbi decoding (recognition) with the updated Lexicon, LM, and adapted HMMs. Viterbi is executed to provide unadapted (*Salida Original*) and adapted (*Salida Adaptada*) word outputs for the spoken sentences. Additionally, the speech's waveform is plotted. The word ouput of the adapted system is then given to a speech synthesizer, which "reads" these words with a more intelligible voice. For this purpose we accessed to the Windows XP *Speech Application Programming Interface (SAPI) ver 5.0*, and the voice used for synthesis was *Isabel* from ScanSoft for spanish.

Another function of this module is to allow dynamic adaptation for the user's HMMs. This was implemented as additional option to the adding of vocabulary in *Añadir Nuevas Frases o Palabras*. Any text written in that form is a stimuli candidate to be read, recorded, and added to the user's personal adaptation speech library. If the user wishes to use any text for adaptation he/she must press the button *Grabar para Adaptación* located under the form. This button works as the other speech recording buttons. Internally, each recording is associated to the stimuli text written in the form and there is no limit about the number of words that can be added. When the user (or the therapist) considers that enough samples have been recorded, he/she just needs to press the button *OK* to perform re-adaptation with all the accumulated speech samples from the user (also the Lexicon and LM are updated).

PERFORMANCE TESTS

Initially the Adaptation and Recognition Modules were tested with two users with normal speech (a female and a male students). The vocabulary for the test consisted of 12 sentences used for control of a robot. Each user read ten times each sentence, thus, 120 sentences were uttered by each speaker. Only whole-sentences were considered, and the ASR performance was 96.67% for the male user, and 94.17% for the female user. As the recognition % of correct sentences was over 94% for both speakers, it can be assumed that % of word recognition accuracy was significantly higher. For the experiments with dysarthric speech, the metric used to measure the performance of the system was % Word Recognition Accuracy (%Acc) which is computed as $\frac{N-D-S-I}{N}$, where D, S, and I are deletion, substitution, and insertion errors in the recognised speech (text output of the ASR module). N is the number of words in the correct ASR's output.

Experiments with Dysarthric Speech

The authorities of the local DIF center provided the support to search and recruit a volunteer to participate in this work. During the search process some requirements were established in accordance to the recommendations of the center's therapists, defining the following: (1) retaining of cognitive understanding; (2) no diagnosis of language understanding impairment; (3) over 15 years old (younger participants require special supervision); and (4) professional assessment of dysarthria.

After a period of two months we could get collaboration from a participant that fulfilled the requirements specified above, GJ. In table 1 the clinical profile of GJ is shown.

Table 1 .

Clinical profile of the dysarthric speaker GJ.

Name:GJ	Age:64	Genre:Male
Pathologies:	low - moderate dysarthria caused by a stroke; left hemiplegia (paralysis of the left arm, leg and trunk); 90% loss of sight; scoliosis (spine is curved from side to side)	

Two ASR systems were tested by this speaker: (1) Speaker-Independent (SI) system, and (2) the Speaker-Dependent (SD) system. As discussed in section *Training Speech Corpus*, this was done to test both approaches as these have been explored in other projects. The **SD ASR System** was tested by GJ without the Speaker Adaptation Module as the system was trained with his speech. To test the SI system, initially GJ had to pass through the Speaker Adaptation Module before using the Speech Recognition Module. The SI system was tested with different amounts of adaptation data to study the effect of static and dynamic adaptation on ASR performance. Three configurations were considered:

- **SI ASR System I:** Base SI ASR system adapted with only the 16 sentences of the Speaker Adaptation Module (static adaptation).
- **SI ASR System II:** SI ASR System I adapted with 11 additional sentences while using the Speech Recognition Module (dynamic adaptation I).
- **SI ASR System III:** SI ASR System II adapted with 11 additional sentences while using the Speech Recognition Module (dynamic adaptation II).

The adaptation sentences for dynamic adaptation were spontaneous sentences related to GJ's activities at his home. These were added to the system's lexicon and LM prior to the test sessions. Finally, all system's configurations were tested with 50 spontaneous sentences (different from those used for static and dynamic adaptation). The results of the test sessions are presented in figure 7. A GSF of 20 was used. The performance of the interface is compared to the performance of other systems: human (HMN) and commercial (CML) ASR on normal speech [14]; commercial SI [3] (CML-Dys) and special purpose SD (SD-Dys) [6] on dysarthric speech.

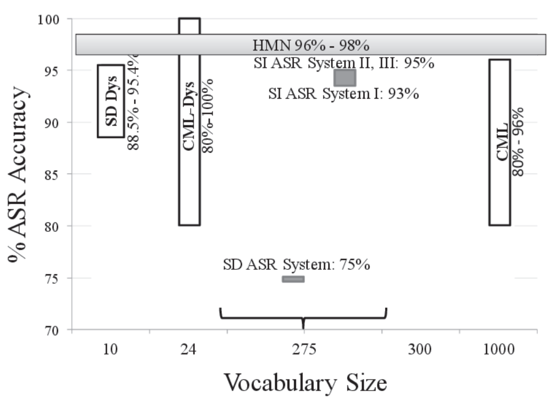

Figure 7 . %Accuracy of the developed ASR interface and comparison with other systems.

As presented, the SI ASR systems had a performance of 93%-95%, achieving the maximum after just one dynamic adaptation. This performance is comparable to human transcription (96%-98%) and commercial ASR for normal speech. When compared with systems adapted (or developed) for dysarthric speakers, this interface achieved performance comparable to those for small vocabularies (< 100 words). Hence, the interface was robust for larger vocabularies (275 words in the test sessions). However the SD ASR performed poorly, with 75% of ASR accuracy.

CONCLUSION

In this paper our advances towards the development of a communication interface for dysarthric Mexican Spanish speakers were presented. We found that a single-speaker speech training corpus, well phonetically balanced, can be used as approach to develop a multi-user ASR system for normal and dysarthric speech. This system, by manipulation of three main variables: gaussian components, vocabulary-LM (control of LM's perplexity), and SGF, can achieve performances up to 95% for dysarthric speech.

The results obtained give confidence about the feasibility of the project and the levels of performance that the system can achive in real-time use. However more research is needed and as future work we have the following:

- to evaluate the performance of the SI ASR system for larger vocabularies (near 1000 words) and dynamic adaptation;
- to analyse the effect of perplexity control and dynamic adaptation for more severe dysarthric speakers;
- to develop an assessment interface to visualize deficiencies in articulation of phonemes.

REFERENCES

[1] Raghavendra, P.et.al. (2001). An investigation of different degrees of dysarthric speech as input to speaker adaptive and speaker dependent recognition systems. *Augmentative and Alternative Communication* 17:265-275.

[2] Moriana, M.J. (2009). La disartria. *In DEP. LEGAL 2922-2007.* ISSN 1988-6047.

[3] Ferrier, L. J.et.al (1995). Dysarthric speaker's intelligibility and speech characteristics in relation to computer speech recognition. *Augmentative and Alternative Communication* 11:165-175.

[4] Jayaram, G., Abdelhamied, K. (1995). Experiments in dysarthric speech recognition using artificial neural networks. *J. of Rehab. Research and Development* 42:162-169.

[5] Strik, H.et.al. (2002). Automatic recognition of dutch dysarthric speech: a pilot study. *ICSLP* 661-664.

[6] Green, P.et.al (2007). A speech-controlled environmental control system for people with severe dysarthria. *Medical Engineering* & Physics 29:586-593.

[7] Copalcua, M.L. (2009). Sistema de procesamiento de fonemas para la rehabilitación de personas con problemas de habla mediante técnicas de aprendizaje automático. *MSc. Dissertation, Technological Institute of Apizaco.*

[8] Pineda, L.A.et.al. (2010). The corpus dimex100: Transcription and evaluation. *Language Resources and Evaluation* 44:347-370.

[9] Young, S., Woodland, P. (2006). The HTK Book (for HTK Version 3.4). *Cambridge University*.

[10] Green, P.et.al. (2003). STARDUSTSpeech Training and Recognition for Dysarthric Users of Assistive Technology. *Association for the Advancement of Assistive Technology in Europe (AAATE)* .

[11] Green, P.et.al. (2003). An Integrated Toolkit Deploying Speech Technology for Computer Based Speech Training with Application to Dysarthric Speakers. In Proc. of Eurospeech 2003, 2213-2216.

[12] Chen, S.et.al. (1998). Evaluation metrics for language models. DARPA Broadcast News Transcription and Understanding Workshop .

[13] Caballero-Morales, S. O. (2011). Automatic Recognition of Dysarthric Speech. Lambert Academic Publishing.

[14] National Institute of Standards and Technology (NIST) (2011). The History of Automatic Speech Recognition Evaluations at NIST. http://www.itl.nist.gov/iad/mig/publications/ASRhistory/ index.html.