

# Research and Development of Speech Technology for Mexican Spanish & Applications for Language training at the Tlatoa Group

**Ingrid Kirschning**

Grupo TLATOA (CENTIA)

Universidad de las Américas-Puebla

Sta. Catarina Mártir, Cholula, Pue., MEXICO

+52 (2) 229-2623, 229-2029

[ingrid@mail.udlap.mx](mailto:ingrid@mail.udlap.mx)

## **ABSTRACT**

Thanks to the advances in today's technology in terms of processing speed of computers, storage space and the management of sound and video devices, speech technology is a reality in almost any kind of computerized system. Speech applications are being used in personal computers, cellular phones, etc. This makes this interesting technology accessible to almost anyone. Among its most useful applications we can find telephone-based information services, banking and computer assisted language learning systems.

There exist already a large number of commercial products that use speech interfaces, developed mainly for English, German and Japanese. That is why we at TLATOA have focused our efforts on making this technology available in the Spanish spoken in Mexico.

To this effect we perform basic research in the different speech processing techniques, trying to improve the performance of speech recognition and synthesis (artificial neural networks, hidden Markov Models (HMM's), Unit Selection, etc.), as well as, the Spanish language, dialogue structure, perception and human-computer interaction approaches, for the development of speech applications.

## **Keywords**

Automatic Speech Processing, Mexican Spanish, Recognition, Synthesis, NLP, Second Language Learning

## **DEVELOPMENT OF SPEECH TECHNOLOGY FOR MEXICAN SPANISH**

TLATOA is founded in 1997 as a result of a joint project with the CSLU (Center for Spoken Language Understanding) of the Oregon Graduate Institute (Portland, Oregon) sponsored by NSF and CONACyT (grant # CO-66-A9605). The CSLU provided us with their Speech Toolkit, thus giving us a good starting point. During the last 2 years TLATOA has grown thanks to the

support of CONACyT, REDII and several collaboration agreements with CSLR, MIT, and companies, like SpeechWorks Intl. The group is becoming larger with students from undergraduate and graduate level, all working on a specific thesis project linked to one of the projects with other institutions or sponsored by the CONACyT.

Specifically, TLATOA has been working in the following areas of automatic speech processing: speech recognition, synthesis, natural language processing, application development, speech technology in education and corpus development. The following sections describe briefly each of these areas.

## **SPEECH RECOGNITION**

Research focuses on the improvement of the methods for training and development of speech recognizers, as well as testing in order to be able to compare their performance with other recognizers [1, 7]. We are now in the process of training a new general-purpose recognizer based on a different neural network architecture as the one used in the CSLU Toolkit. We expect to improve the training time and recognition rate.

## **SPEECH SYNTHESIS**

TLATOA created the first Mexican Spanish voice for Festival, which was later, improved by adding a duration estimation module [4]. Now we are starting to develop a new voice with a better quality, using a Unit Selection approach for the synthesis.

## **NATURAL LANGUAGE PROCESSING**

Based on CSLU's robust parser we integrated natural language processing into query systems that allowed students to obtain information about their courses via a speech interface [8]. The next step is the research on information retrieval systems using mixed initiative for applications in Mexican Spanish.

### APPLICATION DEVELOPMENT

Several small applications have been developed as demos like a voice mail, a system to access e-mail via the telephone among others [5, 6]. Additionally we collaborated with SpeechWorks Intl. In the development of an auto attendant for the university, as web as a system that allows the students check on their account status with the university, also via the telephone.

### SPEECH TECHNOLOGY IN EDUCATION

One very interesting application field for speech technology is the education. As part of a larger project to be developed we created a couple of tools for a computer-assisted language learning environment. These tools are a first prototype for pronunciation verification and a bilingual dictionary, both for Spanish language students, whose native tongue is American English [2,3]. Another very interesting aspect is the use of speech technology based systems to support language acquisition for deaf children. We are developing a system based on the CSLU Toolkit as a first prototype for the Jean Piaget Special Education School in Puebla.

The main challenge here is to create a tool that can be used also by the children that do not speak yet (modulating correctly) and design the interface and the

content of the lessons in a way that is easy to use for the students and easy to manage by the teachers.

### CORPUS DEVELOPMENT

TLATOA has recorded a corpus consisting of 550 adult speakers, mainly from the central area, speaking a large variety of words (names, numbers, digits, letters) as well as spontaneous speech. This has been one of the most time-consuming tasks, but now this corpus, completely transcribed and labeled, is available to anyone without cost for educational and research purposes.

More recently, we have also collected a corpus of almost 1000 children voices ranging from 2<sup>nd</sup> to 5<sup>th</sup> grades. This corpus is being transcribed and labeled now.

### CONCLUSION

TLATOA's focus is set on the development of technology that is accessible to Mexican people and also applications that provide a support/aid to the real needs of Mexican society. One large area where support is needed is in education and language training/acquisition.

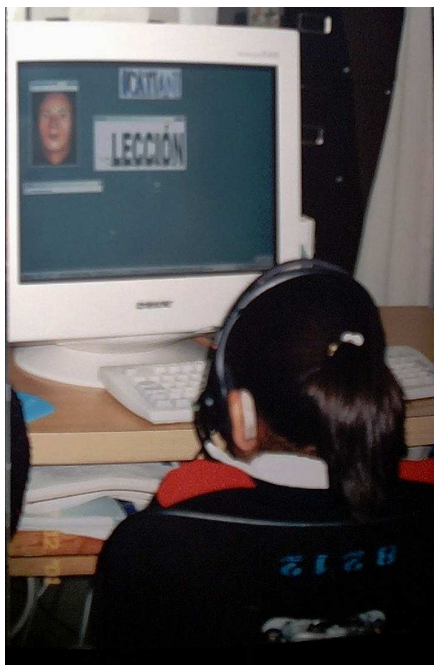


Figure 1: Students from the Jean Piaget School experiment with the system.

## REFERENCES

1. Espinosa, M., Serridge, B. Comparación entre redes neuronales y modelos ocultos de Markov para el reconocimiento de voz, utilizando el CSLU Toolkit, in *Proceedings of ENC'99* (Pachuca, Mexico, September 1999).
2. Kirschning, I., Aguas, N. Verification of Correct Pronunciation of Mexican Spanish using Speech Technology, in *Proceedings of MICAI 2000: Advances in Artificial Intelligence, Mexican International Conference on Artificial Intelligence*, Springer Verlag, (México, April 2000), 493-502.
3. Kirschning, I., Aguas, N. and Ahuactzin, A. Aplicación de Tecnología de Voz en la Enseñanza del Español, in *Proceedings of the 1er. Taller Internacional de Tratamiento del Habla, Procesamiento de Voz y el Lenguaje HAVOL 2000*, (Mexico, July 2000).
4. Meza, H. Modelos Estadísticos de Duración de los Fonemas en un Corpus de Español Mexicano, in *Proceedings of CONIELECOM 2000* (Cholula, Mexico, March 2000).
5. Munive, N., Cervantes, O. Un Sistema de Correo Electrónico y de Voz usando Reconocimiento de Voz, *Soluciones Avanzadas*, 7, 69, (May 1999), 44-48.
6. Munive, N., Vargas, A., Serridge, B., Cervantes O., and Kirschning, I. Entrenamiento de un reconocedor Fonético de Dígitos para el Español de México usando el CSLU Toolkit, *Computación y Sistemas*, 3, 2, (1999), 98-104.
7. Oliver, M.A., and Kirschning, I. Evaluación de métodos de determinación automáticos de una transcripción fonética, in *Proceedings of ENC'99* (Pachuca, Mexico, September 1999).
8. Rosas, O. Sistema de Consultas utilizando Reconocimiento de Voz y Procesamiento de Lenguaje Natural. Master Thesis, Dept. Computer Systems Engineering, UDLAP, (June 1999).