CSLU Toolkit-based Vocabulary Tutors for the Jean Piaget Special Education School

Ingrid Kirschning, PhD
TLATOA Speech Processing Group, CENTIA
Universidad de las Américas, Puebla
Cholula, Pue., México, 72820
ingrid@mail.udlap.mx

Abstract

The TLATOA Speech Processing Group at UDLA, Puebla has developed a couple of applications using the CSLU Toolkit to support language acquisition for hearing impaired children of the Jean Piaget Special Education School in Puebla, México. These applications guide the practice of the pronunciation of vowels and a set of simple short words. The present paper describes the application for vowel production, the test plan and the first results as well as experiences of applying speech technology to support language acquisition for children in a public school in Mexico.

1 Language Acquisition and Hearing Disability

Hearing is the means by which humans acquire their spoken language (Flores & Berruecos, 1991) and (McAleer, 1995). It is through language that a child learns about the surrounding world, about himself and intellectual development takes place.

The hearing impairment can range from a partial to a complete loss of hearing. Hearing impairment or deafness in infants poses a big problem for language acquisition and requires special attention. Although sign language exists it is always preferable that a child, if capable, learns to speak. This is extremely important as the large majority of people including their own parents, teachers and friends may not be able to read or use sign language. This leads to situations where these children are excluded from social groups and miss a large part of their social learning experiences. Therefore it is very important to find the means for these children to learn to express themselves orally (Delgado, 1999).

1.1 Language Therapy and Special Education

In México there are few special education schools and specialized personnel (speech therapists, etc.). We began working together with one particular school, known as a "Multiple Attention Center" which is an institute for children with hearing and language disorders due to different problems, including Down syndrome, trauma, illness, etc.

This school has around 250 students and a staff of 30 teachers which include special learning teachers, psychologists and therapists. The curriculum is dictated by the Mexican Secretary of Public Education or "SEP".

Language therapy is given individually, but there are only few therapists for all the students of the school, and very few can afford private sessions (Valadéz & Espinosa, 2001).

By instructions of SEP the learning method is focused on verbal communication and lip reading. According to (Perelló & Tortosa, 1978) the aim of this method is that the thinking process of a hearing impaired becomes identical to the not impaired, being able to manifest his/her feelings, in spite of articulatory imperfections

Language therapy at this school focuses on three main topics:

- Articulation of vowels and consonants.
- Intonation, modulation and rhythm.
- Individual therapy for oralization

Additionally the kids use a workbook that contains phonemes and pictures to practice. They use mirrors, confetti, posters that illustrate situations, concepts, phrases with intonation marks, colored cards which are used to define gender and the point in time. Most of the material is created by the teachers themselves to support a specific task the class has to do.

1.2 Computer Assisted Language Therapy

There are several programs and applications in the market for language therapy, especially for English language, and one or two for Spanish (Spain). These programs usually focus on one type of therapy: articulation, repetition, intonation or rhythm. Some of the exercises can be used independent of a language (practice of the control of volume, pitch and modulation).

Other cases where correct pronunciation and vocabulary are important the system has to be

developed for a specific language. None of these systems have been developed for Mexican Spanish.

The main concern of the instructors of the school is to be able to personalize the attention to the specific problems of each student, which is nearly impossible due to the lack of personnel. It is thus that an automated tool to support individual language practice was designed. This tool is called ICATIANI and uses speech technology and computer graphics to support pronunciation practice of Mexican Spanish (Kirschning, 2003).

Thanks to the TELETON the Jean Piaget School received 11 Pentium 3 PCs fully equipped, enabling teachers to use software tools for their classes and therapies and opened the possibility to develop new systems tailored to their requirements.

3 Designing the Lessons

Based on a large number of interviews and discussions with teachers and the school's therapist it was decided that the first step was to work on vowel pairs, indicating points and manner of articulation. The system should target the children between ages 3 to 8 years. It should keep personalized registry of each student, their practice sessions, success rate and errors when pronouncing vowels. The system is not intended to substitute a teacher or therapist, it will only be a tool to support their work and give them a little more time to focus on specific problems of each kid while the others can practice with the system on their own.

3.2 Lesson Content

As mentioned before, the method to be used is the oral method, focusing on the explanations of the points of articulation and using lip-reading, no sign language.

The content to be taught in this case is only vowels, as they are the easiest (Delgado, 1999). The five vowels in Spanish are grouped in pairs: /a/ with /o/, /u/ with /e/ and /o/ with /i/, based on similar tongue positions, using their representation in capitals and small caps letters (see fig.1).

3.3 Feedback

The system was developed using the CSLU Toolkit (Cole et al, 1998), (Stone, 1999) and (de Villiers, 2002), which includes 3D animated faces, one of them called Baldi. The face shows the pronunciation of each vowel to be practiced (figures 2a and b). The same face provides the feedback on the correct and wrong answers of the child, smiling if the answer was correct and otherwise making a sad face, explaining that the pronunciation was wrong (figures 3a and b).

The same face that gives instructions is used to give feedback in order to avoid the children being distracted by other images appearing on the screen, such as medals, smileys, etc.

3.4 Program Structure

The system identifies 2 types of user: the instructor and the student. The instructor can register new students, review the results of each student in their personal log-files and also modify lessons. The student logs into the system with his/her full name, then selects a lesson. Baldi will say the instructions for each step, ask for the student to speak and then give feedback of the result of each utterance being recognized correctly or not. The recognizers and the male voice for Mexican Spanish were developed by TLATOA (Kirschning, 2001).

4 ICATIANI

Using CSLU Toolkit's RAD we programmed the logic sequence of the lessons which run in three phases: I) Student login, after which the student is greeted with his name by Baldi.

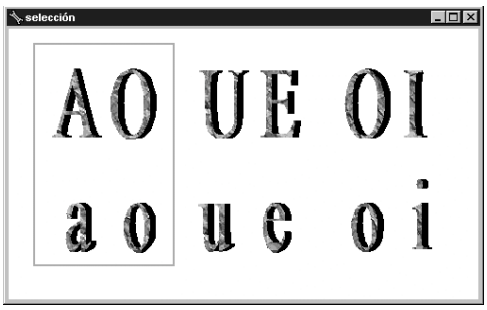

Figure 1: The system lets the student choose a pair of vowels to practice.

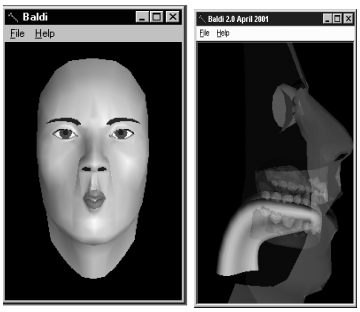

Figure 2: Baldi pronouncing /u/, a) frontal, solid view, b) lateral, semi-transparent view.

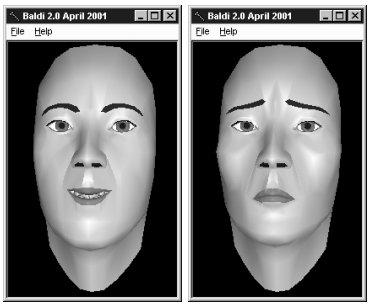

Figure 3: a) Baldi smiling; b) and with a sad face.

This enables the creation of log-files for each student during their whole work session. The log-files contain the date of the session, the practiced vowels and the systems recognition result including the number of trials for each vowel.

II) Presentation, where the student chooses a pair of vowels to practice (fig. 1) by clicking on it, and out of the chosen pair he selects one. Baldi then pronunces the chosen vowel (fig. 2).

III) Practice: after repeating the vowel 3 times Baldi asks the kid to say the vowel. If the recognizer matched the utterance to the correct vowel, Baldi will congratulate the student and smile (fig. 3a), otherwise he will say that it was not the correct sound and ask him to try again (fig. 3b).

4.1 Speech Recognition

It is difficult to use speech recognition with hearing impaired children, because their utterances can sometimes be interpreted correctly by a human listener, even if the pronunciation is not correct, but an automated system will determine the result by using a fixed threshold. Negative recognition results can also result in discouragement of the learners.

Even still, the teachers of Jean Piaget School decided to use the recognizer. They spent the first sessions with each student, explaining why the results were not often good and this surprisingly motivated the kids to try harder.

We tested several settings for different recognizers for children and adults, and an adult speech recognizer was better suited for the deaf children's voices, where the pitch is generally not as high as with hearing kids.

4.2 Baldi

The animated face has a pretty accurate lip movement suited for lipreading, only in some cases the tongue movement is not accurate for Spanish. The speech synthesis is also used even if most of the student cannot hear it. But they sometimes place their hand on the speakers to feel the vibration of the sound when Baldi speaks.

4.3 Captioning

To support the understanding of what Baldi says we added the captioning window. The students that can read, can confirm what they understood by re-reading what was said it in this window.

5 Preliminary Test Results

First, at the laboratory, functionality and usability tests were preformed with 3 children of different ages and abilities to read lips. Here issues like the speed of Baldi's speech were adjusted and the maximum number of repetitions for each vowel.

Figure 4: Students of the Jean Piaget School testing Icatiani together with their parents.

Then another test was made at the Jean Piaget School, where parents tried it with their children, approving the system's use for sessions after school hours (fig.4).

5.1 Test-Plan for Icatiani at Jean Piaget

Assessment Question: "What is the change, if any, in students' vowel production after using the Toolkit for speech practice for six months of the regular school year? Does vowel production come closer to canonical form?"

The previously described system is for vowel practice. Based on the same approach we created an automated version of the kind of test for vowels that teachers use now. It records the kid's speech and the results obtained by using a recognizer. Later the human listeners added their evaluation. Apart form the 55 kids that use Icatiani, a group of 19 kids were be selected as control group with similar etiology to the kids who attend after-school program.

The data of the intelligibility tests will be used to calculate frequencies and descriptive statistics for correctly identified vowels making correlations among naïve listeners. By the end of 2004 we expect to have further test results to analyze and compare the kid's speech production after training with Icatiani, to determine if significant change occurred over time.

5.2 Initial Assessment

The initial test was performed with 55 students of 4 different groups. The chart in figure 5 shows the intelligibility rate for each vowel per group. Vowels /e/ and /o/ are those with best results, and /u/ and /i/ have the lowest rates overall with the recognizer. When evaluated by a human listener

(fig. 6) these rates increase drastically because, we believe, often pronunciation was close, barely audible but still enough for the evaluators to guess the correct vowel.

Based on these results we decided to incorporate a volume meter that shows in a colored scale if the utterances are within an acceptable volume range. It helps the children realize how loud they are speaking, and learn to control their voices better.

6 Conclusions

The reason for using the CSLU Toolkit is because it is probably the simplest and easiest to incorporate speech tools and educational software in standalone special purpose applications.

The tools developed with the CSLU Toolkit for English language acquisition have been evaluated and the results published recently, encouraging us to continue in this pursuit. Barker states in her paper, that the results indicate that the CSLU vocabulary tutors may be an efficient way to increase vocabulary and literacy in oral-deaf education (Barker, 2003).

Icatiani has been developed with a similar approach to that of the CSLU Vocabulary Tutors, only that Icatiani includes also speech recognition. Speech recognition creates a high rate of rejections, and several researchers consider it too disturbing for the children. However, so far practice with these children at this particular school has proven the contrary.

The initial assessment has shown that the level of oral production of the children is extremely low and they require more than Icatiani, but Baldi seems to motivate them to produce speech, and teachers have said that even if production is not improving fast, they are less shy to try and speak in class.

The desired assessments have not been finished as some changes in the schools staff and policies made us loose precious time. In the meantime, thanks to Icatiani groups of 11 children at a time can practice their vowel production, and every time Baldi smiles it evokes a big smile in the children, motivating us to continue this work.

Presently new additions are being made to Icatiani to practice diphthongs and simple words.

References

L. Barker. 2003. Computer Assisted Vocabulary Acquisition: The CSLU Vocabulary Tutor in Oral-Deaf Education. *Journal of Deaf Studies and Deaf Education,* 8:2, Oxford University Press.

R. Cole, T. Carmell, P. Connors, M. Macon, J. Wouters, J. deVillers, A. Tarachow, D. Massaro, M. Cohen, J. Beskow, J. Yang, U. Meier, A. Waibel, P. Stone, G. Fortier, A. Davis, C. Soland. 1998. Intelligent Animated Agents for Interactive Language Training. STiLL: ESCA Workshop on Speech Technology in Language Learning, Stockholm, Sweden.

R. Cole. 1999. Tools for research and education in speech science. Proc. of the 14th International Congress of Phonetic Sciences; San Francisco, USA.

J.J. Delgado. 1999. Hipoacusia infantil. PrevInfad, June 14.

L. Flores & P. Berruecos. 1991. El niño sordo de edad preescolar. Ed. Trillas, México.

I. Kirschning. 2001. Research and Development of Speech Technology & Applications for Mexican Spanish at the Tlatoa Group. Development Consortium at CHI 2001, Seattle,WA, pp.49-50.

I. Kirschning. 2003. Icatiani: A system to support language acquisition for profoundly deaf children. VII Mexican Symposium on Medical Physics, UNAM, México.

P. McAleer Hamaguchi. 1995. Childhood Speech, Language & Listening Problems. John Wiley & Sons, U.S.A.

J. Perelló, F. Tortosa. 1978. Sordomudez. Ed. Científico-Médica, Audiofoniatría y Logopedia, Barcelona, España, 3rd. Ed.

P. Stone. 1999. Revolutionizing language use in oral deaf education. Proc. 14th International Congress of Phonetic Sciences; San Francisco, U.S.A.

Valadéz, & M.R. Espinosa. 2001. Personal communication, Jean Piaget School.

J. de Villiers. 2002. CSLU Vocabulary Tutor. http://www.cslu.ogi.edu/toolkit/docs/2.0/apps/vocabulary/.

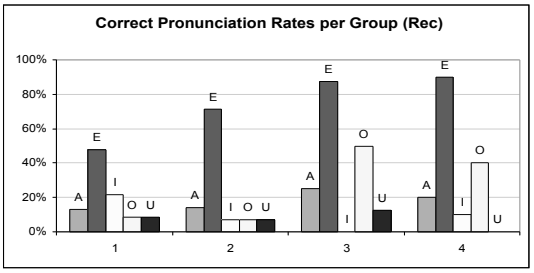

Figure 5: Intelligibility rates (speech recognizer).

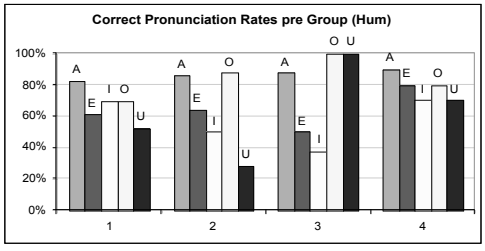

Figure 6: Intelligibility (human listener).