ISCA Archive
http://www.isca-speech.org/archive

Speech and Language Technology in Education (SLaTE2009)
Wroxall Abbey Estate, Warwickshire, UK
September 3-5, 2009

# An Experience with a Spanish Second Language Learning Tool in a Multilingual Environment

*Oscar Saz*[1], *Victoria Rodríguez*[2], *Eduardo Lleida*[1], *W.-R. Rodríguez*[1], *C. Vaquero*[1]

[1]Communications Technology Group (GTC)
Aragón Institute for Engineering Research (I3A), University of Zaragoza, Zaragoza, Spain
[2] Vienna International School, Vienna, Austria
oskarsaz@unizar.es,vrodriguez@vis.ac.at,{lleida,wricardo,cvaquero}@unizar.es

## Abstract

This paper presents the results of an experience with "VocalizaL2", an application for Second Language (L2) learning of Spanish, in a multilingual environment at the Vienna International School (VIS). For the experiment, a group of 6th-graders at the school practiced with the application during 5 sessions altogether with their regular classes. The results of the experiment show on one hand, the great motivation power that computer-based L2 tools have for the pronunciation training of young learners, while also resulting useful for the teachers. On the technical aspect, the tool and the algorithms within are described and a preliminary analysis points out their ability to correct and motivate non-native Spanish pronunciation.

## 1. Introduction

Achieving a correct pronunciation of a language is an important issue in the learning process of that language. Pronunciation skills are necessary to be able to interact and communicate properly with native and non-native speakers in the new language. In traditional theories, pronunciation training was mostly neglected from the activities of second language teachers where the approaches were based in studying the vocabulary, syntax and grammar of the new language, considering that the pronunciation would be learned instinctively by the student. However, more recent theories have given back to pronunciation training the relevance it has in oral communication [1].

Possibilities of applying speech technologies to pronunciation training in Second Language (L2) learning and Computer-Aided Language Learning (CALL) have increased dramatically in the recent years with the development of several tools for languages like English [2], Chinese [3], Japanese [4], Dutch [5] and others. Novel algorithms to detect word and phoneme mispronunciations are developed to support these tools and provide correct and robust feedback to the new students. However, a language like Spanish, with more than 400 million native speakers around the world, is lacking further development of L2 learning tools to promote and improve the knowledge of Spanish pronunciation to students interested in it.

This work explores the possibility of integrating novel advances in pronunciation verification algorithms within already deployed tools for speech therapy and education in Spanish. This exploratory research [6] was performed in a multilingual environment at the Vienna International School (VIS) and tries to understand how students and teachers react to new software applications while trying to make an initial evaluation of the performance of the pronunciation verification algorithms used.

This work was supported by national project TIN2008-06856-C05-04 from the Spanish government

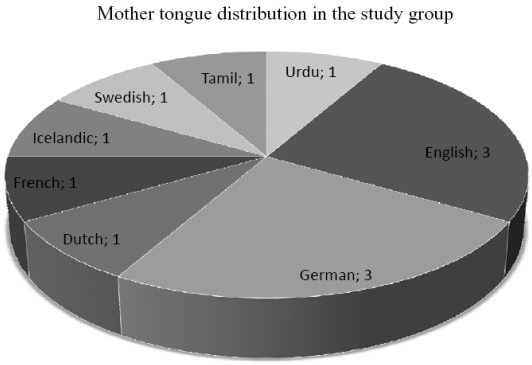


Figure 1: *Distribution of mother tongues in the experimental group*

The paper is organized as follows: Section 2 will describe the environment in which the learning experiment was run, while Section 3 will present "VocalizaL2", the application used for the experiment and the algorithms within the application. The outcome of the experience is explained in Section 4. Finally, Section 5 will provide the discussion and conclusions to this work.

## 2. Description of the School Environment

The experiment described in this work was performed at the Vienna International School (VIS). VIS gathers children, among others, from officials of the different United Nations agencies located in the city of Vienna and diplomatic delegations. Language teaching is one of the pillars of the multicultural education in VIS; with English and German as the official languages in the School, students also have access to study their mother tongue and a third language (French or Spanish) of their choice.

The present work was made during a period of 5 non-consecutive weeks, where a 45-minute session per week with the application was run weekly. During each session, every student could practice with the application for a time of approximately 10 minutes in one of the 2 computers in which the application was installed. The experimental group of students was composed of 12 students in 6th grade (11 years old). In this grade, students start the learning of the third language, so all of them were beginners in Spanish. In the group there were 8 boys and 4 girls and the distribution in terms of the mother tongue of every student is seen on Figure 1. Mother tongue and the

Table 1: *Words per session*

| Session | Subject | Words |
|---|---|---|
| 1 | Food | beber, bocadillo, botella, carne cereales, cerezas, chocolate galletas, hamburguesa, helado |
| 2 | Daily routine | andar, bañarse, cocinar, dormir ducharse, escribir, jugar llorar, pegar, trabajar |
| 3 | Animals | araña, ardilla, burro, cerdo conejo, foca, gallina mariposa, pájaro, rinoceronte |
| 4 | At home | armario, cocina, cuadro, bañera escalera, frigorífico, sillón ventana, espejo, librería |
| 5 | Neighborhood | acera, ambulancia, balcón, calle calzada, farola, papelera |

interlingua that students use to interact with the new language (classes are taught on English) are an important issue to understand how different students can react to different pronunciation learning issues.

The set of words was chosen according to the vocabulary studied by the children in the classes and divided among the 5 sessions as on Table 1. Ten words were finally chosen for each session, and each student pronounced 2 times each word. Session 5 had a different approach and was designed as a gameplay in which students were divided in teams and they pronounced alternatively the words; only 7 words were, hence, programmed for this session.

The application, "VocalizaL2", offers the possibility of giving pictorial, auditive and written prompts to the user. The three options were chosen during sessions 1,2,3 and 5. In session 4, the written prompt was omitted to observe how students dealt with only auditive prompting.

## 3. VocalizaL2

The application used for the experimental trial was "VocalizaL2", this application gathered all the experience in education tools for children under "Comunica" [7] and the research in phoneme-level pronunciation verification [8] with a recently acquired corpus containing speech from handicapped children who produced multiple mispronunciations [9]. While "Vocaliza" [10] was aimed to provide with corrective speech therapy to handicapped children, "VocalizaL2" aimed to provide with a more precise feedback to students of higher educational needs or learners of Spanish as a second language.

### 3.1. Pronunciation verification algorithm

The pronunciation verification algorithm within "VocalizaL2" was based on the diagram in Figure 2 and was similar to works for pronunciation verification systems tested in tasks of assessment of impaired speech [8]. Hidden Markov Models (HMM) used for the forced alignment and in the phone network were trained from Spanish adult speech corpora; 25 acoustic units representing the 23 phonemes of the Spanish language and the glides /j/ and /w/ were represented in the model, where each unit was a 3-state HMM and every state had a distribution as a Gaussian Mixture Model of 16 Gaussians. A Mel-Frequency Cepstral Coefficient (MFCC) feature extractor was used that used 12 cepstral coefficients plus the first and second derivatives plus the log-energy.

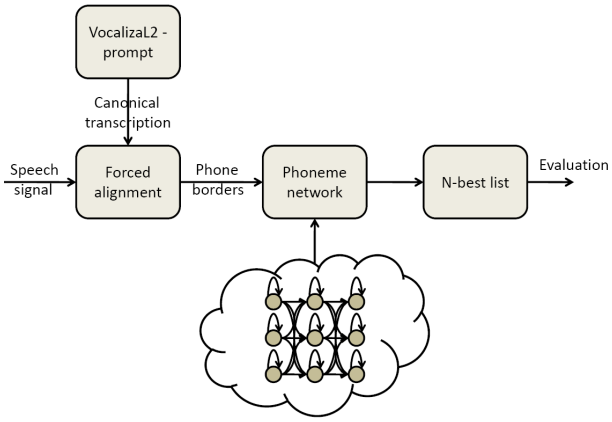


Figure 2: *Pronunciation verification diagram*

In a first stage, a Viterbi-based forced alignment estimated the phoneme borders according to the canonical transcription of the prompted word in the input speech signal. Posteriorly, an unconstrained phoneme network was applied, where all phonemes were studied in all possible positions within the word. An N-best phonetic network was hence obtained for each word position with the log-likelihood score obtained by each phoneme. From the N-best network, the position of the phoneme that matched the canonical transcription in each position was evaluated. When the canonical phoneme obtained the first position in the N-best list, the evaluation given to the phoneme was a positive value as the difference between the log-likelihoods of the canonical phoneme and the phoneme that ranked the second position in the N-best list. If the canonical phoneme did not obtain the first position, the evaluation was a negative value marking the difference between the log-likelihoods of the phoneme ranked in the first position of the list and the canonical phoneme.

### 3.2. User interface

User interface of "VocalizaL2" was the same as in the original version of "Vocaliza" [10]. The use of Alternative and Augmentative Communication (AAC) systems, designed initially for sensory handicapped users, was also used to attract and keep the interest of children like the target group of this experience. Only difference was in the presentation of the results to the speaker; more information was shown in the new application where the graphemes of the prompted word were shown in screen with different colors according to the final result given to the associated phoneme as on Figure 3. Colors ranged from green (good pronunciation) to yellow (average pronunciation) and red (poor pronunciation). With only three levels, the users (young children) could have a more straightforward way of understanding the quality of their pronunciation. An overall evaluation value (displayed at the bar in the right side) was provided as the average result of all the phonemes in the word.

## 4. Results of the Experience

The experience proposed in this work had two different aspects to evaluate. On one hand, the pedagogical results of using a computer-based tool for L2 learning in the proposed environment; and, on the other hand, the evaluation of the tool and its ability to correctly verify the pronunciation of the users. Both aspects are presented in this Section.

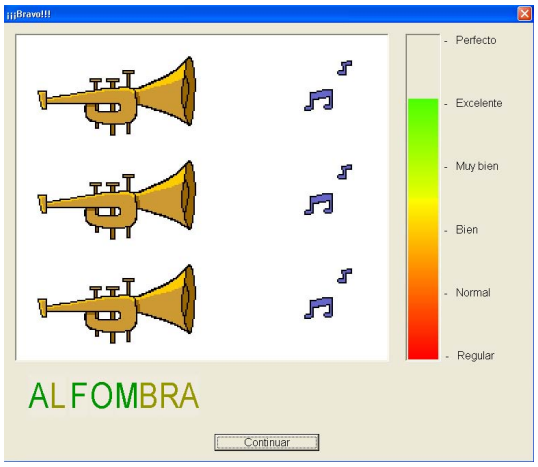


Figure 3: *Pronunciation feedback*

### 4.1. Pedagogical results of the experience

The opinions of the students with the application were collected afterwards each session by their teacher. Students were aware, despite their short age, that the application was providing an evaluation coherent with the effort and interest they put on their pronunciation; so they strongly tried to improve in different trials and sessions. They evaluated positively the interface and felt really motivated to have more classes with the application, although they also pointed out some weak points like the lack of naturalness of the synthetic voice that provided the audio prompt and the sometimes odd evaluation results given by the tool (possibly due to the presence of noises or disfluences).

Considering the different possibilities provided by the application to L2 teachers, the teacher realized that session 4 without the written prompt was more challenging for students, as they had to rely only on the audio prompt, which could be more interesting for advanced students. Session 5 was seen positively by all students and teacher, as gameplay activities provided an extra motivation to young learners.

### 4.2. Evaluation of the tool

Initial approach was to make an evaluation of the performance of the tool reviewing the evaluation results given by the tool to all speakers. Unfortunately, no labeled data was available from the speakers and the short number of sessions made difficult a more precise study of the results. Hence, the performance evaluation was made studying the log-likelihood scores that the application assigned to each utterance and that were kept stored within the application. Three points of relevant interest were studied: Evaluation in different trials of the same word, trend of the evaluation through different sessions, and specific results for different phonemes and words.

#### *4.2.1. Evaluation in different trials*

Two different trials of every word were programmed in the sessions. Average results obtained by the students in the first and second trials of every word are provided on Figure 4. A review on the results separating both trials showed that evaluation marks given on the second trial were higher than in the first trial. This was consistent with the fact reported by the teacher that students put a bigger effort in the second trial after they got the evaluation of the first trial; furthermore re-prompting for the second trial reinforced the correct pronunciation of the word.

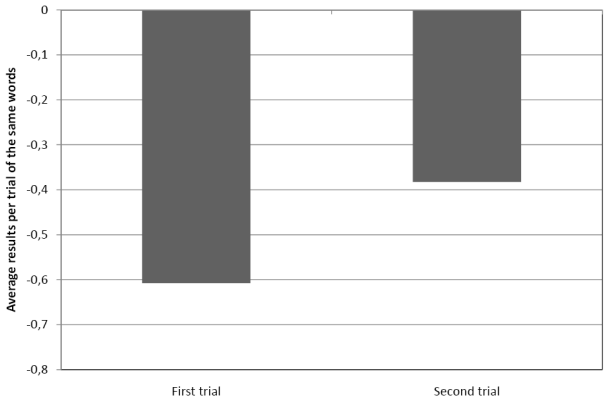


Figure 4: *Average evaluation results across different trials*

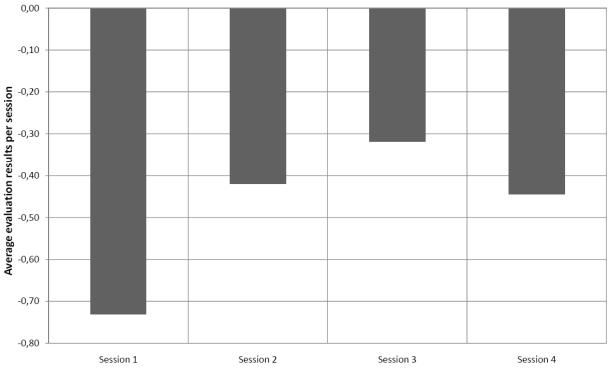


Figure 5: *Average evaluation results across sessions*

#### *4.2.2. Evaluation in different sessions*

Also, the average results obtained by the students in each session are shown in Figure 5. These results showed an important improvement in the evaluation obtained from session 1 to session 2, followed by a lesser improvement in session 3 and a reduction in the results in session 4. Improvement from session 1 to session 2 could be explained by the fact that the children got used to they way in which application worked and they understood it better putting more effort in their pronunciation. Results in session 4 (slightly worse than previous session), might corroborate the fact pointed out by children and teacher that uttering words without the written prompt was harder for students.

#### *4.2.3. Evaluation for different phonemes and words*

Finally, a study of the performance of the ability of all speakers to pronounce different words and phonemes was performed. The average log-likelihood results (according to the method in Section 3.1) for all the 25 units used (23 phonemes plus allophones $/j/$ and $/w/$) for all speakers and sessions are shown on Table 6. Lack of labeled data diminished the ability to extract conclusions, but trends were similar to the trends of mispronunciations in young Spanish children in the natural process of language acquisition [11]. Students achieved higher marks on vowels, while special phonemes like $/tS/$, $/L/$ or $/J/$ received lower marks. Also noticeable it is the significant worst results achieved in the glides $/j/$ and $/w/$ compared to the corresponding vowels $/i/$ and $/u/$; even if the sound is similar, it is usually noticed how glides in diphthongs are more difficult to pronounce than the vowels for young Spanish children [11].

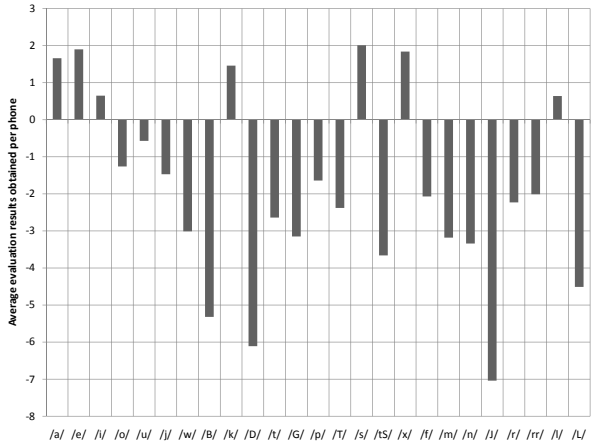


Figure 6: *Average evaluation results per phoneme*

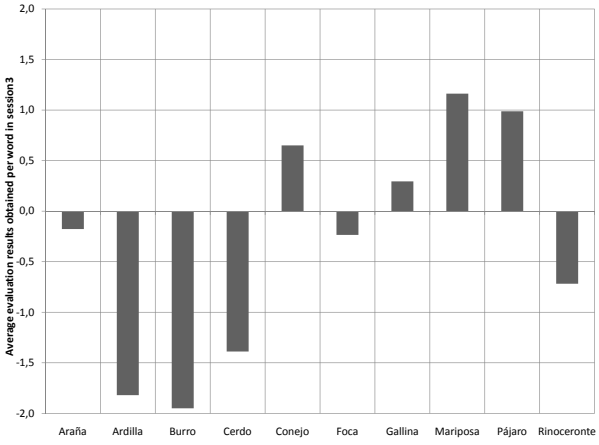


Figure 7: *Average evaluation results across words in session 3*

A study over the results given to a set of words was also made, the 10 words studied were the words in session 3 as it was the session in which the students were more used to the application. The average results for all speakers with these words are shown on Figure 7. Again, lack of labeled data made unable a comparative study, but some trends could be observed that again agreed with theories of phonetic acquisition in Spanish. Words $burro$ (SAMPA: $/Burro/$), $ardilla$ (SAMPA: $/arDiLa/$), $cerdo$ (SAMPA: $/TerDo/$) and $rinoceronte$ (SAMPA: $/rrinoTeronte/$) were the worst pronounced by the students, which was consistent with the fact that these words contained the phoneme /rr/ or the phoneme /r/ in coda position, which is usually a difficult feature for learners of Spanish.

## 5. Conclusions

As conclusion to this work, an extension of the tools within the "Comunica" framework to L2 Learning has been presented. This tool has been evaluated in a multilingual environment with children enrolled in classes for learning Spanish. Students and teacher have reviewed the great motivational power of computer-based tools; and the ability of the algorithms within has been reported to the extend that the lack of labeled data allowed.

Further work arises as result of this experimental work. In terms of user interface, a major adaptation has to be done to adapt the application environment to an interface more suitable for adults learners. In terms of the pronunciation evaluation method within the application, the introduction of phonotactic rules to constrain the phoneme network used to decode the most probable phoneme sequence is strongly required. These rules would limit the arcs in the network and would avoid impossible phoneme sequences that might introduce some noise in the evaluation process.

## 6. References

[1] J. Morley, *Pronunciation pedagogy and theory: New view, new directions*. Alexandria (VA), USA: TESOL Publications, 1994.

[2] J. Lee and S. Seneff, "Automatic grammar correction for second-language learners," in *Proceedings of the 2006 International Conference on Spoken Language Processing (ICSLP - Interspeech)*, Pittsburgh (PA), USA, September 2006.

[3] G. Fengpei, P. Fuping, L. Changliang, D. Bin, and Y. Yonghong, "Forward optimal modeling of acoustic confusions in mandarin call system," in *Proceedings of the 2008 International Conference on Spoken Language Processing (ICSLP - Interspeech)*, Brisbane, Australia, September 2008, pp. 2807–2810.

[4] C. Tsurutani, Y. Yamauchi, N. Minematsu, D. Luo, K. Maruyama, and K. Hirose, "Development of a program for self assessment of Japanese pronunciation by English learners," in *Proceedings of the 2006 International Conference on Spoken Language Processing (ICSLP - Interspeech)*, Pittsburgh (PA), USA, September 2006.

[5] A. Neri, C. Cucchiarini, and H. Strik, "Improving segmental quality in l2 dutch by means of computer assisted pronunciation training with automatic speech recognition," in *Proceedings of CALL 2006*, Antwerp, Belgium, 2006, pp. 144–151.

[6] V. Rodríguez, "El uso de herramientas multimedia para la práctica de la pronunciación en clases de ele con adolescentes," Memoria final del Máster en Enseñanza del Español como Lengua Extranjera (MEELE), 2008, Universidad Antonio de Nebrija, Departamento de Lenguas Aplicadas.

[7] W.-R. Rodríguez, O. Saz, E. Lleida, C. Vaquero, and A. Escartín, "Comunica - tools for speech and language therapy," in *Proceedings of the 2008 Workshop on Children, Computer and Interaction*, Chania, Greece, October 2008.

[8] S.-C. Yin, R. Rose, O. Saz, and E. Lleida, "Verifying pronunciation accuracy from speakers with neuromuscular disorders," in *Proceedings of the 10th International Conference on Spoken Language Processing (ICSLP-Interspeech)*, Brisbane, Australia, September 2008, pp. 2218–2221.

[9] O. Saz, W.-R. Rodríguez, E. Lleida, and C. Vaquero, "A novel corpus of children's impaired speech," in *Proceedings of the 2008 Workshop on Children, Computer and Interaction*, Chania, Greece, October 2008.

[10] 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, pp. 4509–4512.

[11] L. Bosch-Galcerán, *Evaluación Fonólogica del Habla Infantil*. Barcelona, Spain: Ed. Masson, 2004.