

# The use of CALL in acquiring foreign language pronunciation and prosody – General specifications for Euronounce Project

N. Cylwik,\* G. Demenko,\* O. Jokisch,\*\* R. Jäckel, M. Rusko,\*\*\*  
R. Hoffmann,\*\* A. Ronzhin,† D. Hirschfeld,‡  
U. Koloska,‡ and L. Hanisch§

\*Adam Mickiewicz University, Institute of Linguistics, Poznan, Poland

\*\*TU Dresden, Laboratory of Acoustics and Speech Communication,  
Dresden, Germany

\*\*\*Slovak Academy of Sciences, Institute of Informatics, Bratislava, Slovakia

†Russian Academy of Sciences, Institute of Informatics and Automation,  
Petersburg, Russia

‡voice INTER connect GmbH, Research and Development, Dresden, Germany

§REZO Computer-Service GmbH & Co. KG, Dresden, Germany

nataliac@amu.edu.pl

lin@amu.edu.pl

## ABSTRACT

The paper presents technical, linguistic and didactic specifications for Euronounce project which aims at creating an intelligent tutoring system with multimodal feedback functions for acquiring foreign languages' pronunciation and prosody. In response to the European Union's call for promoting less widely spoken languages the project focuses on German as a target language for speakers whose mother tongue is Polish, Slovak, Czech or Russian and vice versa – it is to enable German native speakers to acquire the pronunciation and prosody of Polish, Slovak, Czech or Russian. Beside specifications concerning corpora design, speech databases, recordings, structure of exercises and feedback system the article outlines theoretical underpinning of the project as well as the baseline for the project, AzAR, created in two preceding projects.

## 1. Introduction

Technological development the world has been facing during last decades enabled technological advance also in such areas as education. The increasing use of technology can be especially seen in the area of foreign language learning which has led to the establishment of a new discipline known under the name of Computer-assisted language learning (CALL). CALL systems have evolved together with the changing approach to the teaching of foreign languages, i.e. the focus has shifted from teaching writing skills, grammar and vocabulary to teaching oral skills and thus also pronunciation and prosody [8]. Teachers and researchers in the field have insisted on paying more attention to segmentals and suprasegmentals the main argument being their importance

for communication [1, 5]. However, there are also other motives for acquiring L2 pronunciation and prosody: as mentioned by Jokisch et al. [7] strong foreign accent may cause integration problems which makes it particularly important in the times of global migration and the policy of integration. The growing interest in teaching and learning foreign language pronunciation and prosody has been reflected in the development of Computer-assisted pronunciation learning (CAPT) within which programs based on different technologies have been created.

## **2. State of the art**

The beginnings of CALL were mostly limited to PC-based exercises to gradually become more and more interactive using graphs, animation, audio and video elements [8]. Modern CALL systems are widely exploring new techniques to help users learn foreign languages such as the use of speech analysis and recognition, animated agents, talking heads and virtual tutors which are to serve mainly feedback purposes. Although yet a few years ago it was often doubted whether speech recognition systems were advanced enough to be integrated in tutoring software [8, 10] there have been more and more positive results reported recently, e.g. for correcting segmental and durational errors [3, 5].

## **3. Advantages of CALL**

The literature on CALL mentions a number of its potential advantages: elimination of time limitations [3, 10] and dependence on the teacher, i.e. the user can study wherever, whenever and as much as it is suitable for him/her, the possibility to work at the learner's own tempo, the possibility to store the user's profile to monitor the progress [10], access to a number of additional material such as visualizations, recordings, animations [6], individualization of the learning process [10] and elimination of the stress related to the fact that the learner is being listened to by his/her colleagues [3], the last of which seems particularly important in the case of pronunciation/prosody learning.

Although it is often mentioned that there is still no strong evidence on the long-term effectiveness of CALL [1, 8], students' positive attitude towards technology use in L2 learning has been reported, which can certainly support the learning process [8].

The overview of the literature on CALL lets us state that "It is clear that the benefits of CALL have been widely accepted, and educators agree that it can be an effective instructional tool" and therefore "at present, the focus is not on whether to accept computer technology. Rather, research is now centered on how to integrate technology more effectively into the learning and teaching of languages" [8].

## **4. Euronounce project**

Intelligent Language Tutoring System with Multimodal Feedback Functions (acronym Euronounce) is a project within the framework of European Commission's Lifelong Learning Programme which aims at creating L2 pronunciation and prosody teaching

software. In accordance with EU's policy of promoting less widely spoken languages the project focuses on Slavonic-German language pairs: Polish-German (PL/DE), Slovak-German (SK/DE), Russian-German (RU/DE) and Czech-German (CZ/DE). The Euronounce project was preceded by two earlier projects carried out by Euronounce coordinator, TU Dresden, between 2004 and 2007. As a result an audio-visual software AzAR (German acronym for Automat for Accent Reduction) aimed at teaching Russians German pronunciation was created [7]. Following the baseline developed in these projects the Euronounce project aims at creating software for pairs L1 DE – L2 RU, CZ, SK, PL and L1 RU, CZ, SK, PL – L2 DE beside segmental adding also supra-segmental exercises.

#### **4.1. Speech databases and speaker selection**

It seems clear that in order for pronunciation tutors to be successful not only target but also source language needs to be taken into account [3, 6]. It is understandable if we keep in mind that most errors result from L1 and L2 interference and consist primarily in transferring allophonic and phonotactic rules from our mother tongue to the target language and replacing L2 phonemes with their most similar L1 counterparts [9]. Taking only L2 into account is one of the main flaws of ASR-based pronunciation tutors as they mostly fail to recognize non-native speech [7, 10]. This is why in the development of Euronounce software 3 speech databases are to be created for each language pair:

- A. reference database – target language read speech by target language native speakers,
- B. non-native speech database – target language speech by non-native speakers, reflecting typical pronunciation and prosody mistakes in the target language,
- C. source-language accent database – source language speech by source language native speakers for ASR training and comparative study of interferences.

The exact structure of the databases for the pair combination L1 Polish, L2 German is showed in the table 1 below.

For B speech database it is planned to record speakers at different levels according to Common European Framework of Reference for Languages [2], i.e. levels A1-A2, B1-B2, C1-C2. For each level at least 6 speakers will be recorded. Two possible methods of speakers' recording are being taken into consideration so that they reflect different levels: recording the same speakers making progress over a certain period of time or recording different speakers at different levels. The speakers will be university students, males and females, possibly subjected to the same educational conditions, i.e. the same course, teacher, educational material. For each subject meta-data will be gathered and stored in the form of a survey on language courses taken, other foreign languages spoken, time spent abroad, certificates gained, etc. It is not intended that each speaker is recorded for the whole B corpus since it is obvious that elementary students will not be able to perform some tasks, e.g. tasks for spontaneous speech production.

For A speech database it is planned to record part of the same material as for B database, read by target language native speakers. This way, it will be possible to compare native and non-native pronunciation and prosody for the same speech material. Corpus C will include about 100 hours of the source language native speech.

**Table 1. The structure of the databases for the pair L1 Polish – L2 German**

Database	Language used	L1 of the speaker	Speech material
A	German	German	Dialectological standard test (Veith test) Phonetically balanced and rich sentences Texts for fluent reading
B	German	Polish	Dialectological standard test (Veith test) – 100 short sentences, 40 complex sentences Phonetically balanced and rich sentences Accent test – special material devised for specific language pair on the basis of pre-assessment of potential errors Texts for fluent reading, retelling and argumentation – approx. 6 between 200–500 words Tasks for spontaneous speech production Interview
C	Polish	Polish	Phonetically rich and balanced sentences – approx. 300 sentences and 10 continuous speech passages

## 4.2. Recordings

The recordings will be conducted with the use of WiGE software in a quiet room or studio with low noise and reverberation, using 2-channel input, i.e. close-talk and table/condenser microphone. Basic quality requirements are: sampling frequency 44,1 kHz, minimal resolution 16 bit, minimal SNR of 35 dB.

## 4.3. Linguistic structure of test material

The literature on CALL emphasizes that pedagogical theory and basic rules of L2 learning should be implemented in educational software [5, 8, 10]. Zinovjeva [10] claims that in order for it to be possible teachers should get acquainted with speech technology. However, due to authoring systems it has become possible for teachers and phoneticians to transfer their knowledge into tutoring systems without being experts on speech technology [7]. Such authoring system was already created within AzAR projects, which allowed for a vast cooperation of linguists, phoneticians, experts on speech technology and philologists. As a result extensive linguistic curricula are being prepared by teachers, philologists and phoneticians for practicing each language pair.

Exercises are being designed in the way that allows practicing production as well as perception at the phonemic and prosodic level in isolated words, simple phrases, complex phrases and continuous speech.

### 4.3.1. Segmental features

Test structure will consist in material for production, perception and discrimination of L2 sounds in minimal pairs, in contrast with L1 sounds and in larger syntactic units in

order to practice assimilations within and between words and phrases as well as missing/inserted syllables, words and phrases.

#### *4.3.2. Suprasegmental features*

Exercises will be devised in order to test and practice prosody in smaller and larger syntactic units. In isolated words suprasegmental identification will be devoted mainly to the perception and production of regular and irregular lexical stress and feet structure as well as types of nuclear accents, duration, intensity, identification of mono-, di-, tri-, four-syllable words. At the level of simple and complex sentences exercises will consist in production and recognition of different types of sentences, i.e. declaratives, commands, wh-questions, yes/no questions, compounds, requests on the basis of their suprasegmental features. Also identification and production of emphatic stress, relating focus with meaning and performing communicative functions with focus will be practiced and tested e.g. showing emotions, disagreement, correcting wrong information, calling attention to new information.

#### **4.4. Feedback system**

Lack of proper (or any) feedback is often named as the most serious flaw in educational software [1, 4, 6, 7]. Good software should not only assess if the pronunciation is correct or incorrect but also instruct on how to improve it, show where exactly the error has been made, e.g. which phone has been produced erroneously [4, 7] and offer feedback that is easy to interpret [1, 10]. The aim of the Euronounce project is to answer these needs through creating a system that would provide multimodal feedback. The baseline for this system was already developed in the AzAR projects. As a result, the AzAR software includes visual and audio modules in the form of curriculum recordings by a reference voice and the visualization of the speech signal under the transcribed and phonemically segmented reference utterances. The software uses HMM-based speech recognition and speech signal analysis on the learner's input due to which the user can visually and aurally compare his/her own performance with that of the reference voice. Most importantly, the system also includes automatic error detection on the phonemic level. All uttered phones are marked using color scale from red for mispronounced phones to green for those pronounced correctly. In other words, the user can listen to and play back the model voice as well as see the speech signal for a particular utterance, record and listen to his/her own utterance and see the speech signal for his/her own utterance and finally get feedback on his/her own pronunciation. Additional visual mode includes animated visualization of the vocal tract (lips area and articulators movements) and a formants graph for particular phones. A typical AzAR template for an exemplary phrase is showed in Figure 1.

This type of design allows adjusting the software to the user's strategy of learning and can make the process of acquiring L2 pronunciation and prosody more efficient. Positive results of audio-visual feedback have been reported especially in the context of prosody teaching [1, 4, 10] whereas for segmental practice also traditional instruction is being recommended since visuals can be too difficult for the user to interpret and listening drill is not enough when one keeps in mind that L2 learner tends to associate foreign sounds with more familiar L1 sounds [4]. Therefore, beside audio-visual feedback, AzAR software includes also text tutorial on articulatory and basic acoustic

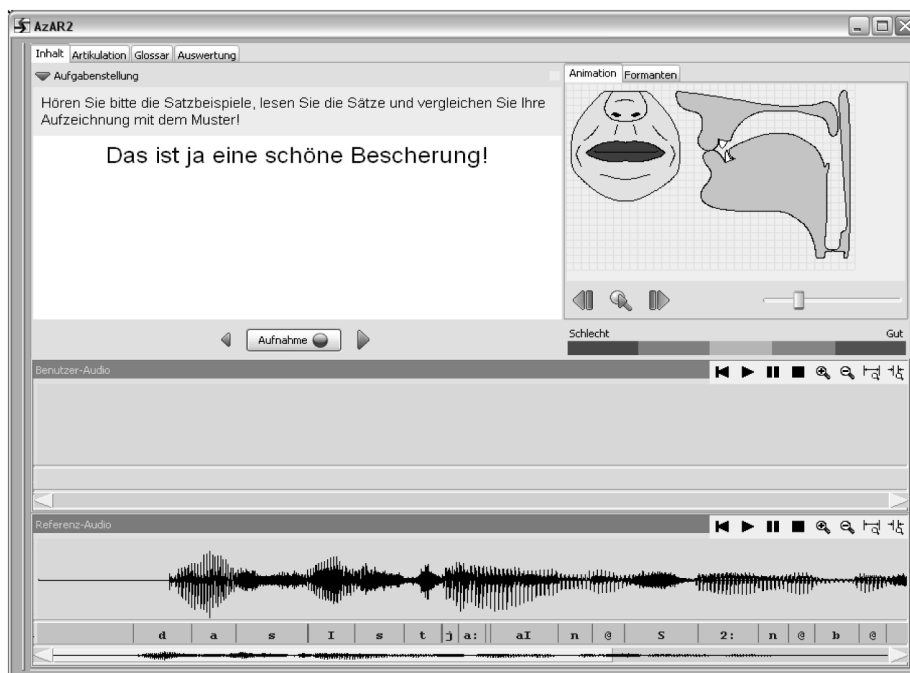


Figure 1. AzAR template for the pronunciation assessment of an exemplary phrase “Das ist ja eine schöne Bescherung”.

phonetics with glossary, phonemes description and classification, anatomic information, etc.

## 5. Conclusions

This paper presents linguistic, technical as well as didactic specifications for the Euronounce project which aims at creating an intelligent system with multimodal feedback functions for learning pronunciation and prosody of German, Polish, Slovak, Czech and Russian. The article outlines basic theoretical foundations as well as the core technology established in the preceding projects based on a German-Russian language pair. This baseline coupled with new cross-lingual databases are to help improve the visualization and quality assessment methods and to allow including prosodic factor in the final software.

**Acknowledgements and disclaimer.** This project has been funded with support from the European Commission within the Lifelong Learning Programme (project 135379-LLP-1-2007-1-DE-KA2-KA2MP). This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein. The project homepage is located at: <http://www.euronounce.net>.



Education and Culture DG

Lifelong Learning Programme



## BIBLIOGRAPHY

- [1] Chun, D.M. 1998. Signal analysis software for teaching discourse intonation. *Language Learning & Technology*, 2(1), pp. 61–77. Retrieved on 15th July 2008 from <http://lt.msu.edu/vol2num1/article4/>.
- [2] Common European Framework of Reference for Languages. Retrieved on 21st August 2008 from [http://www.coe.int/t/dg4/linguistic/illustrations\\_EN.asp](http://www.coe.int/t/dg4/linguistic/illustrations_EN.asp).
- [3] Eskenazi, M. 1996. Detection of foreign speakers' pronunciation errors for second language training – preliminary results. *Proc. of the Intern. Conf. on Spoken Language Processing (ICSLP)*, Philadelphia, PA., 1996. Retrieved on 15th July 2008 from <http://www.asel.udel.edu/icslp/cdrom/vol3/096/a096.pdf>.
- [4] Eskenazi, M. 1999. Using automatic speech processing for foreign language pronunciation tutoring: some issues and a prototype. *Language Learning & Technology*, 2(2), 1999, pp. 62–76. Retrieved on 15th July 2008 from <http://lt.msu.edu/vol2num2/article3/index.html>.
- [5] Eskenazi, M., Hansma, S. 1998. The Fluency pronunciation trainer. *Proc. Speech Technology in Language Learning*, Marholmen, 1998. Retrieved on 15th July 2008 from [http://www.cs.cmu.edu/~max/mainpage\\_files/Esk-Hans-98.pdf](http://www.cs.cmu.edu/~max/mainpage_files/Esk-Hans-98.pdf).
- [6] Engwall, O., Wik, P., Beskow, J., Granström, G. 2004. Design strategies for a virtual language tutor. In Kim, S. H. & Y. (Eds.), *Proc. of the Intern. Conf. on Spoken Language Processing (ICSLP)*, Jeju Island, 2004, pp. 1693–1696. Retrieved on 16th July 2008 from [http://www.speech.kth.se/~olov/list\\_of\\_publications.html](http://www.speech.kth.se/~olov/list_of_publications.html).
- [7] Jokisch, O., Koloska, U., Hirschfeld, D. and Hoffmann, R. 2005. Pronunciation learning and foreign accent reduction by an audiovisual feedback system. *Proc. of 1st Intern. Conf. on Affective Computing and Intelligent Interaction (ACII)*, Beijing, 2005, pp. 419–425.
- [8] Liu, M., Moore, Z., Graham, L., Lee, S. 2002. A Look at the Research on Computer-Based Technology Use in Second Language Learning: A Review of the Literature from 1990–2000. *Journal of Research on Technology in Education*, 34(3), pp. 250–273.
- [9] Wells, J.C. 2000. Overcoming phonetic interference. *English Phonetics, Journal of the English Phonetic Society of Japan*, 3, pp. 9–21. Retrieved on 20th March 2008 from <http://www.phon.ucl.ac.uk/home/wells/interference.htm>.
- [10] Zinovjeva, N. 2005. Use of speech technology in learning to speak a foreign language. Retrieved on 16th July 2008 from [http://www.speech.kth.se/~rolf/NGSLT/gslt\\_papers\\_2005/Natalia2005.pdf](http://www.speech.kth.se/~rolf/NGSLT/gslt_papers_2005/Natalia2005.pdf).

