

Phonetic and Prosodic Aspects in the Cross-lingual Pronunciation Tutoring

Oliver Jokisch

Chair for System Theory and Speech Technology
Dresden University of Technology, 01062 Dresden, Germany
oliver.jokisch@tu-dresden.de
<http://www.ias.et.tu-dresden.de>

Abstract. Computer-assisted pronunciation tutoring (CAPT) methods have been well-established in research and education. Common system approaches include the phonetic quality assessment, highlight problematic sections in the speech signal and usually rely on automatic speech recognition (ASR) regarding the target language L2. The contribution deals with the audiovisual CAPT system AzAR. An extensive feedback mechanism and several speech databases for Slavonic learners of German were developed. Currently, AzAR is adapted for Chinese Learners of German and for learners of the Basque language. This extended abstract summarizes some experiences of the multilingual experiments and system tests – focusing on phonetic and prosodic assessment aspects.

Keywords: Pronunciation tutoring (CAPT), Slavonic languages, Mandarin Chinese, German, Basque

1 Slavonic-German Transfers in Pronunciation Tutoring

The speech data collection and the adaptation of automatic speech recognition (ASR) algorithms are essential tasks for the development of pronunciation tutoring (CAPT) systems. The system "Automat for Accent Reduction (AzAR)" [1] was originally designed for the pronunciation tutoring of learners with a native language L1 from the Slavonic language group and for the target language L2 German. Within the cooperation project Euronounce [2], the concept was extended to the Slavonic target languages Polish, Slovak, Czech and Russian. The Euronounce database includes special lessons for phonetic peculiarities but also sentences to evaluate the prosodic aspects. It contains 130 speakers and about 200 hours of speech. In further steps, the Euronounce concept was also tested for Mandarin Chinese learners of German (as L2 and L3) and learners of the Basque language with different mother tongues. Preliminary results with focus on phonetic and prosodic aspects are summarized in the following sections.

2 Cross-lingual effects in Chinese learners of German

In this section, some experiences from the AzAR adaptation for Chinese Learners of German [3–5] are reported.

2.1 Analysis of phonetic and prosodic deviations

With acoustic and perceptual investigations, it was found that various segmental and intonational deviations contribute to difficulties in oral communication:

- Inaccurate production of those German vowels and consonants which are nonexistent in Chinese,
- Incorrect placement of tonal categories and wrong phonetic realization of a phonological category.

Chinese learners usually employ different strategies, such as epenthesis, deletion and modification to deal with unfamiliar sounds because Chinese syllables usually end with vowels, and the learners usually add a schwa /@/ after the consonant final. In a comparative analysis of German produced by Russian and Chinese learners, it was obvious that epenthesis occurs more frequently among Chinese learners than Russian learners [4].

With the visual-audio feedback information and after many times of trial and error, the learners became conscious of their pronunciation mistakes, and could make correspondent corrections. Chinese is a tone language, Chinese speakers thus raise or lower their pitches to express different lexical meanings instead of different linguistic purposes in intonation languages like German and English. The f₀ deviations in Chinese speakers of German have been investigated in [3].

The visualization of intonation curve is proved to be particularly effective in the acquisition of L2 intonation. Automatic pitch tracking algorithm, however, usually displays many small pitch changes that make the learners confused about the sentence intonation, and moreover some small changes are linguistically unimportant.

2.2 Preliminary results

It proves possible for Chinese learners of German language to imitate standard pronunciations successfully. However, a faithful imitation of isolated words or sentences with visual aids can not guarantee a good pronunciation in ordinary speech. The articulatory constraints will still dominate for many students in normal speech without any audiovisual aids. The tutoring system should also guide the learners step by step from a successful imitation to an accurate production in free continuous speech. Therefore the next research interest will be focused on continuous and normal speech, which is the ultimate goal of language teaching.

3 Basque language as example of a new learning target

In the following section, the peculiarities of the target language L2 Basque are discussed – basing on the contributions [6, 7].

3.1 Language peculiarities

Basque is an isolated language which does not belong to the Indo-European language group, as one would expect from its geographical location [8]. It has two major languages as neighbors, Spanish and French, and the influence of these languages on Basque is noticeable, especially on people who are learning Basque as L2. Moreover, the fact that the Academy of the Basque Language has not made any decision about standard Basque intonation or prosody yet, may confuse new students regarding their reference in phrase accent and intonation.

Considering the importance of prosody acquisition in Basque, a suprasegmental analysis part has been added to database and CAPT system which will be useful for future developments. In the suprasegmental part, word level and sentence level intonation have been taken into consideration – concerning the segmental part, some phonetic and some phonological features:

- Phonetic features: phonemes that do not exist in the neighbor languages, as the /ts'/ [9], the differentiation between the six sibilants of Basque and the vocalic system.
- Phonological features: the palatalization process in the context /iV/ and /inV/ (V refers to a vowel), and the unvoicing process caused by the negative particle *ez* in the first phoneme of the next word.

A baseline Basque curriculum was designed, and a 16 kHz/16 bit PCM speech database was recorded from a Basque native speaker. The segmental part of the database consists of 125 sentences and 60 word pairs. The suprasegmental part contains 20 isolated words and 50 sentences.

3.2 Phonetic assessment

The ASR baseline system for the segmental part involves an HMM-based phoneme verification system in forced alignment mode, using GOP (Goodness Of Pronunciation) score as confidence measure. It was trained using a Basque studio database which contains recordings from native and non-native speakers, as well as dialectal and standard Basque data.

Measuring the general performance of the ASR system for the reference voice, it was concluded that the algorithm was not able to discern properly between sibilants. The explanation of this fact is that there are Basque speakers of different skill levels in the training database and so wrongly pronounced utterances were used to train the HMMs of phonemes that do not exist in Spanish. Later, a new data set was trained using only the native speakers, so that more robust decisions were obtained.

3.3 Prosodic assessment

For the suprasegmental part, a simple but efficient approach for the prosodic assessment was tested by directly calculating the RMSE between the realized and the reference f0 curve. For that purpose, a test corpus was recorded with

speakers whose mother tongue is different. They were asked to read some Basque sentences where prosodic segments were indicated, and the meaning of each part was translated to them – to force the application of their native intonation to the Basque sentence.

References

1. Jokisch, O., Koloska, U., Hirschfeld, D., Hoffmann, R.: Pronunciation learning and foreign accent reduction by an audiovisual feedback system. In Proc. 1st Intern. Conf. on Affective Computing and Intelligent Interaction (ACII), Beijing (China), pages 419-425, October 2005. Springer LNCS-3784.
2. Jokisch, O., Jaeckel, R., Rusko, M., Demenko, G., Cylwik, N., Ronzhin, A., Hirschfeld, D., Koloska, U., Hanisch, L., Hoffmann, R.: The EURONOUNCE project - An intelligent language tutoring system with multimodal feedback functions: Roadmap and specification. Proc. Konferenz Elektronische Sprachsignalverarbeitung (ESSV), pages 116-123, September 8-10, 2008. Frankfurt/M.
3. Ding, H., Jokisch, O., Hoffmann, R.: F0 analysis of Chinese accented German speech. Proc. 5th Intern. Symposium on Chinese Spoken Language Processing (ISCSLP), p. 49-56, Dec. 2006. Singapore.
4. Hilbert, A., Mixdorff, H., Ding, H., Pfitzinger, H., Jokisch, O.: Prosodic analysis of German produced by Russian and Chinese learners. Proc. 5th Intern. Conf. on Speech Prosody, May 11-14, 2010. Chicago.
5. Ding, H., Mixdorff, H., Jokisch, O.: Pronunciation of German syllable codas of Mandarin Chinese speakers. Proc. Konferenz Elektronische Sprachsignalverarbeitung (ESSV), 281-287, Sept. 2010. Berlin.
6. Odriozola, I., Jokisch, O., Hernaez, I., Hoffmann, R.: A Pronunciation Tutoring System for Basque - First Development Steps. Proc. Konferenz Elektronische Sprachsignalverarbeitung (ESSV), Aug. 2012. Cottbus.
7. Odriozola, I., Navas, E., Hernaez, I., Sainz, I., Saratxaga, I., Sanchez, J., Erro, D.: Using an ASR database to design a pronunciation evaluation system in Basque. In Proc. 8th Intern. Conf. on Language Resources and Evaluation (LREC), Istanbul, May 2012.
8. Hualde, J. I.: Basque Phonology, Taylor & Francis, 1991 (ISBN 9780415056557).
9. University of Bilbao, Aholab: SAMPA computer readable phonetic alphabet of Basque, retrieved 25 April 2012 from http://aholab.ehu.es/sampa_basque.htm