# CAPT-Systems Azar3 and RehaVOX: Stage of Development and Prospects

*Rainer Jaeckel, Institute of Acoustics and Speech Communication, TU Dresden*

The Institute of Acoustics and Speech Communication of TU Dresden co-operated between 2004 and 2014 with different partners in the development of acoustic and audio-visual feedback systems for use in CAPT/L2-learning and speech therapy. This paper presents the CAPT application Azar3 (2010) and the speech therapy application RehaVOX (2014).

The approach that was chosen in the Azar projects from the beginning on was *individual error detection*, e. g. pronunciation errors are localized in the student's speech input and the feedback module indicates the degree of deviation from a standard L2 speech pattern. Therefore, linguistic units are mapped onto the acoustic speech signal using *forced alignment recognition (FAR)*.
The feedback system utilises the confidence measures of the aligner for speech quality assessment. In the LLP project EURONOUNCE a multilingual tutoring system was created which considers alternative pronunciations, e.g. phoneme variants which belong to the student's L1 or correspond to different levels of L2 acquisition. Alternative pronunciation variants were marked in a significant number of tokens and included in the dictionary of the aligner as alternative transcriptions.

The pronunciation tutoring system Azar3 consists of two major components: 1. learning program as front end module for the language learner; 2. data input and administration module (authoring system), which serves as tool for the language teacher and with the help of which the content to be learned can be captured and administrated. The tutoring system employs forced alignment for individual error detection. The aligner segments the incoming speech signal and labels the segments according to the expected phoneme string. The HMM-based feedback system with native-trained acoustic models relies on confidence measures.

The user interface of Azar3 presents an animated frontal view of the mouth region with the ''skin'' on and a moving sagittal view of the articulatory organs. Other means of visualization are the wave form diagram of the reference signal and the incoming signal (both related to the aligned speech segments), and an interactive formant chart. The wave form diagram can be changed by the spectrogram or F0 contour. Focus can be set on segments or larger units. The play function is synchronized with the animated views of speech organs. In the speech analysis window the incoming signal is visualized as waveform-diagram in the area above the diagram of the reference sample. The feedback system performs segmentation/alignment of the student's record and computes the confidence scores for each segment. According to adjusted thresholds, different colours – varying between dark green and dark red - are assigned to the speech segments. The interactive colour scale is sensitive to changes so that the student gets an immediate and detailed return on segmental accuracy for any iteration of an exercise. The EURONOUNCE tutoring system provides content modules for diverse learning scenarios and different L2 instruction levels (a total of more than 6000 exercises).

In the speech therapy application RehaVOX the reliability of feedback has been increased due to combined implementation of forced alignment recognition and five additional feedback plugins that analyse (1) the f0-contour, (2) the short time spectrum, (3) the intensity contour, (4) the formant structures of vocoids, and (5) the time labels of phone segments and pauses. Acoustic scores, e.g. a Goodness of Pronunciation (GoP) score for each segment, are generated from the output of the feedback plugins. The speech therapy application RehaVOX is designed as a server-client-architecture, where the acoustic analysis functions are executed on the server.

Continuing the work in the fields of L2 pronunciation training, we are drafting an adaptive CAPT-system that will be able to model the acquisition of standard L2 pronunciation by a speaker of a certain L1 as a process, and to support an optimal learning strategy.