propor 2008

International Conference on Computational Processing of Portuguese Language

Applications of Portuguese Speech and Language Technologies

Applications of Portuguese Speech and Language Technologies - Propor 2008 Special Session

Hosted by:

Universidade de Aveiro

Promoted by:

Microsoft Language Development Center

Propor 2008 Special Session Commitee

Special Session Chair

- **António Teixeira -** DETI/IEETA, Universidade de Aveiro, Portugal

Organising Committee

- **Daniela Braga,** Microsoft Language Development Center, Portugal
- **Miguel Sales Dias,** Microsoft Language Development Center, Portugal
- **António Teixeira -** DETI/IEETA, Universidade de Aveiro, Portugal

Programme Committee

- **António Teixeira -** DETI/IEETA, Universidade de Aveiro, Portugal
- **Daniela Braga,** Microsoft Language Development Center, Portugal
- **Vera Strube de Lima,** Pontifícia Universidade Católica do Rio Grande do Sul, Brasil
- **Luís Caldas de Oliveira,** INESC-ID/IST, Portugal

Editorial Board

- **Daniela Braga,** Microsoft Language Development Center, Portugal
- **Miguel Sales Dias,** Microsoft Language Development Center, Portugal
- **Luanda Braga Batista,** Microsoft Language Development Center, Portugal

EUROPEAN PORTUGUESE VOICE TECHNOLOGY FOR FOREIGN LANGUAGE LEARNING AND FOR ARTICULATION DISABILITY IMPROVEMENT

Luis Coelho, Daniela Braga

Instituto Politénico do Porto, MLDC – Microsoft Language Development Center
lcoelho@eu.ipp.pt, i-dbraga@microsoft.com

Application Description

In this demo an application for articulation improvement and/or language learning is presented. This software, named LIPS-EP (*Language Improvement and Practicing System*), is aimed for children that have some kind of articulation pathology or to those who start to learn the basic vocabulary in a foreign language, in this case European Portuguese (EP), and want to improve their pronunciation skills. The graphical user interface has been developed with nice and appealing images and cartoons and is extremely easy to use. Though it is expected that the user alone can fully use the software in some cases the presence of a speech therapist or a language teacher is recommended for explaining the articulation details. An example screenshot of the user interface is presented in figure 1.

Software usage follows the following sequence: In the main menu the user selects a set of words grouped according to a desired criterion. It is possible to choose environments (like jungle, home, etc.) or group words that have a specific phoneme. The user is then presented with a set of images (within the selected category) and it should pronounce what is shown after clicking one of the images. The speech is recorded and simultaneously a text-to-speech system (TTS) generates the same word. Both speech samples are then compared, matched together and the similarity is calculated. The result is plotted against time along with the spectrogram for each of the speech samples. The similarities are painted using a semaphore like colour code that clearly show what phonemes have been correctly pronounced. The more frequent and experienced users can also look at the details in the spectrograms and to the phonetic sequence. If a phoneme is clicked a schematic head with the articulation organs in the related articulation position will be shown. By listening to the generated speech and after viewing what sounds/phonemes have fewer similarities the user can correct his pronunciation by looking at the correct articulation position.

The comparison mechanism between the user's recorded speech and the synthetic speech can be configured in several ways. The similarity metric can be changed and so the colour thresholds can. A time warping option gives the possibility to perform the comparison using synchronous time or by first performing an alignment of the two utterances. This is an important option that can be used for improving speaking rhythm.

A management interface is also provided for extending the basic set of words originally distributed with the application. With this interface it is possible to add new words to the database by indicating the name, picture and category.

Conclusion and Future Work

The described application is being used by some children and speech therapists with very nice feedback. However some improvements are still foreseen. The main challenge is to develop at least two child voice fonts, male and female, that could have a greater similarity with the user's voice and reduce vocal tract differences. This would create a better spectral reference for voice analysis that could lead to best comparison results.