Reactive accent interpolation through an interactive map application

Maria Astrinaki[1], *Junichi Yamagishi*[2,3], *Simon King*[2], *Nicolas d´Alessandro*[1], *Thierry Dutoit*[1]

[1]Circuit Theory and Signal Processing Lab, University of Mons, Belgium
[2]The Centre for Speech Technology Research, University of Edinburgh, Edinburgh, UK
[3]National Institute of Informatics, Tokyo, Japan

maria.astrinaki@umons.ac.be, jyamagis@inf.ed.ac.uk, Simon.King@ed.ac.uk
nicolas.dalessandro@umons.ac.be, thierry.dutoit@umons.ac.be

Abstract

MAGE enables the reactive and continuous models modification in the HMM-based speech synthesis framework. Here, we present our first prototype system for extended interpolation applied for interactive accent control. Available accent models for American, Canadian and British English are manipulated in realtime by means of a gesturally controlled interactive geographical map. The accent interpolation is applied to one gender at a time, but the user is able to reactive alter between genders, while controlling the speakers to be interpolated at a time.
Index Terms: speech synthesis, reactive, dialect, interpolation

1. Reactive HMM-based speech synthesis

In the application, various English accents need to be controlled and interpolated in realtime. Therefore, we use MAGE[1], which supports a realtime architecture for reactive HMM-based speech synthesis. MAGE uses multiple threads, and each thread can be affected by the user, allowing accurate controls over the different production levels of the artificial speech [1]. Accessing and controlling the thread responsible for the model manipulation we can reactively modify the way the available models are used for the parameter generation. MAGE allows the reactive control of the the interpolation weights of every feature stream for every phonetic label, as illustrated in Figure 1. This feature allows reactive and continuous control over the degree of interpolation between various models, maintaining any other controls.

2. Reactive accent interpolation map

In order to separate out speaker characteristics and accent so that listeners can focus only on accent transitions we use multiple speakers who have similar accents, by interpolating their acoustic models. As users interact with the map application[2] they selected the speakers for interpolation. All speakers are chosen from the CSTR voice banking corpus.

The application consists of the world map, on which every single speaker is represented as a circle. The "active" region controlled by the user is represented as a yellow circle around the cursor. The user can zoom in/out, navigate, select the speaker's gender and the interpolation "mode" by using the standard mouse or touchscreen controls. There are two ways to interpolate between speakers: "collision" mode, where the active region overlaps and selects one or more speaker for interpolation and "continuous" mode, where each time the cursor moves, the distance between the cursor and all the available

[1]MAGE : http://mage.numediart.org/.
[2]A video can be found in https://vimeo.com/67662099.

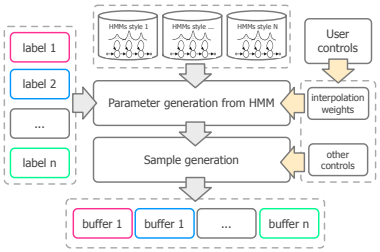

Figure 1: Reactive parameter generation by continuously interpolating multiple models.

speakers is computed and the N-nearest speakers are selected to be interpolated as shown in Figure2. When the voice models are selected, the interpolation weights are computed (uniform weights of $w = 1/N$), the speech parameters are generated and the speech output is synthesized.

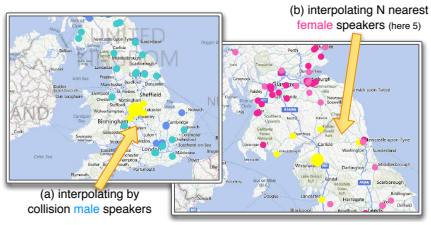

Figure 2: Examples of interactive accent interpolation during (a) "collision" and (b) "continuous" modes.

3. Conclusions

The interactive accent map application can have several applications, targeting the creation of unique personalized voices. In the field of new interfaces for musical expression and performing arts, in gaming, movie dubbing or GPS applications as well as assistive applications for speech impaired people. Finally in speech pedagogy and therapy by creating adaptive references for certain dialects [2]. However it is not straightforward to formally evaluate the proposed interactive accent control.

4. References

[1] M. Astrinaki and N. d´Alessandro and L. Reboursière and Alexis Moinet and T. Dutoit, MAGE 2.00: New Features and its Application in the Development of a Talking Guitar, Proc. of NIME´13.

[2] M. Tachibana, J. Yamagishi, T. Masuko, and T. Kobayashi, Speech Synthesis with Various Emotional Expressions and Speaking Styles by Style Interpolation and Morphing, IEICE Transactions, E88-D, 11, 2484–2491, 2005.