Final Report for SICSA Distinguished Visiting Fellowship

Ricardo Gutierrez-Osuna

Department of Computer Science and Engineering, Texas A&M University

1. SUMMARY OF WORK

The bulk of my work focused on evaluating HMM-based synthesizers for the purpose of foreign accent conversion. Junichi Yamagishi had noted that adapting average (native) voice models to a foreign speaker leads to a reduction in the speaker's perceived foreign accent. Based on this observation, I sought to study whether there was a "sweet spot" in the process, defined as an ideal number of adaptation sentences where the identity of the target speaker is captured while his/her accent is transformed to sound more native. In the process, I implemented objective measures of foreign accent, speaker identity and synthesis quality. The foreign accent measure was based on a correlation of forced-alignment scores across phonemes between synthesized utterances from the foreign speaker and those from the average model. The speaker identity measure was based on a linear discriminant analysis of MFCC and F0 characteristics of the resynthesized utterances. Finally, the synthesis quality was based on P.563, a single-ended measure of speech quality for narrow-band telephony.

I performed the adaptation study using scripts developed by Junichi Jamagishi (HTS2007). Namely, I built adaptation models for 8 male speakers: 3 American English speakers (MLK, BDL, and RMS from Arctic), 2 British speakers (AWB from Arctic, Roger from Blizzard) and 3 non-native speakers (KSP from Arctic, EZW, a native speaker of German, and RGO, a native speaker of Spanish). For each speaker, I built adaptation models starting with one adaptation sentence (in which case only pitch characteristics were adapted) and up to fifty sentences (in which case spectral envelope characteristics were also adapted), then repetitions each.

In addition, I also participated in two recordings of articulatory (EMA) data. One recording consisted of over 800 sentences from the Glasgow Herald corpus produced by a non-native speaker of English (RGO); a second parallel corpus consisted of 400 utterances by a native speaker of American English (MAB).

2. PRESENTATIONS AND VISITS

During my stay at CSTR I prepard four presentations: August 24: CSTR group meeting; October 22: GIST, University of Glasgow; November 24, CSTR Listen! meeting; December 3, IPAB, University of Edinburgh.

On October 22 I visited the University of Glasgow to meet with Steve Brewster, Matthew Chalmers, and Roderick Murray-Smith. During that visit I also discussed applications of our foreign accent conversion with Pascal Belin, whose research interests are in fMRI studies of voice perception. I also contacted John Arnott at Dundee University and expressed interest in visiting with him to discuss collaborations in pervasive health. Unfortunately, the only suitable dates for a visit were during my last week in Edinburgh, at which point my focus was on completing the HTS adaptation study.

3. FUTURE COLLABORATIONS

I continue to collaborate with Junichi Yamagishi on the speaker adaptation study. For this purpose, I have provided copies of all my code (adaptation scripts and objective measures) to Cassia Valentini-Botinhao and am working with her to extend the study beyond 50 adaptation sentences, which turned out to be too few to achieve full speaker adaptation. A second collaboration with Christian Geng (Linguistic) is in the processing of the EMA recordings, which may also lead to collaborations with Korin Richmond. Namely, my interest is in investigating whether articulatory information can be used to effectively transform foreign accents. A third (potential) area of collaboration is with Mirjam Wester, with whom I discussed using our objective measures to analyze recordings from her study of bilingual speakers.