Articulatory Speech Synthesis from Vocal-Tract MRI Data

Rachel Alexander, Asterios Toutios, Shrikanth Narayanan
Signal Analysis and Interpretation Laboratory (SAIL), University of Southern California, USA

Introduction

• Using a model-based method of articulatory synthesis, outlines of the midsagittal vocal tract provide a substantial basis for simulating human speech production.
• A speech synthesizer has been programmed in MATLAB, which, in the long term, will output sound signals corresponding to given inputs of dynamically changing vocal-tract shapes, obtained through real-time magnetic resonance imaging data from several speakers.

Background and Data

• Articulatory synthesis: Synthesis of speech acoustics by simulation of the physics of the propagation of sound in the vocal tract and the dynamics of the vocal-tract shaping. The time-domain simulation was developed by Maeda [1].
• Vocal-tract data: Vocal-tract area function dynamics for vowel-consonant-vowel sequences were generated by interpolating area functions for static vowels and consonants derived from MRI data by Story and Titze [2]. Inputs describing the area of the glottis were derived by rules described by Maeda [3] and Toutios and Maeda [4].

Methodology

• The synthesizer creates a consistent glottal pulse comprised of slow- and fast-varying components, given data representing the amplitudes of each component, the frequency of glottal vibration, and a variable sampling frequency, to simulate the noise source emitted from the vocal cords, using empirical rules developed by Maeda [3].

Results

• The synthesizer generated vowel-consonant-vowel sound sequences, including voiced, voiceless, and nasal consonants, with reasonable intelligibility compared to the original recorded sound.

Conclusion

• Thus far, acoustics of vowel-consonant-vowel sequences have been synthesized on the basis of MRI data. Expansion to synthesis of short sentences is currently under investigation.
• With further developments, this synthesizer could potentially produce fully comprehensive text-to-speech simulation, using available vocal-tract data to generate more realistic and extensive articulatory synthesis results.

References


Email: rachela@usc.edu