Title:
Synthesizing Speech Intra/Inter Speaker and Language

Speaker:
Frank K. Soong, Principal Researcher/Research Manager
Speech Group, Microsoft Research Asia (MSRA)
Beijing, China

Abstract
The phonetic and acoustic nature of a person's speech is strongly conditioned by his own articulators and the language he speaks. It is therefore both academically interesting and technically challenging to investigate how to render speech intra/inter-speaker and language wise. The rendering quality is usually assessed in three criteria: naturalness, intelligibility and speaker similarity. All three criteria is in general not easy to be satisfied altogether when rendering is done cross-speaker and/or cross-language. In this talk we will analyze the key factors which cause quality rendering difficult acoustically and phonetically. Speech databases in the same language but recorded by different speakers or bilingual speech databases recorded by the same speaker(s) are used. Both acoustic and phonetic measures are adopted to quantify naturalness, intelligibility and speaker similarity. Our "trajectory tiling" algorithm-based, cross-lingual TTS is used as the baseline system for inter-language rendering. To equalize speaker difference automatically, DNN-based ASR acoustic model trained speaker independently is used. Kullback-Leibler divergence is proposed to measure the phonetic similarity between any two given speech segments, which can be from different speakers or languages, for selecting good candidates. Demos will be given to show various rendering results either intra/inter speaker or language.

Frank K. Soong,
Principal Researcher/Research Manager
Speech Group, Microsoft Research Asia (MSRA)
frankkps@microsoft.com

Frank K. Soong is a Principal Researcher and Research Manager, Speech Group, Microsoft Research Asia (MSRA), Beijing, China, where he works on fundamental research on speech and its practical applications. His professional research career spans over 30 years, first with Bell Labs, US, then ATR, Japan, before joining MSRA in 2004. At Bell Labs, he worked on stochastic modeling of speech signals, optimal tree-trellis, N-best decoding algorithm, speech analysis and coding, speech and speaker recognition. He was responsible for developing the recognition algorithm which was developed into voice-activated mobile phone products rated by the Mobile Office Magazine (Apr. 1993) as the "outstandingly the best". He is a co-recipient of the Bell Labs President Gold Award for developing the Bell Labs Automatic Speech Recognition (BLASR) system. He has served as a member of the Speech and Language Technical Committee, IEEE Signal Processing Society and other society functions, including Associate Editor of the IEEE Speech and Audio Transactions and chairing IEEE Workshop. He published extensively with more than 250 papers and co-edited a widely used reference book, Automatic Speech and Speech Recognition - Advanced Topics, Kluwer, 1996. He is a visiting professor of the Chinese University of Hong Kong (CUHK) and a few other top-rated universities in China. He is also the co-Director of the National MSRA-CUHK Joint Research Lab. He got his BS, MS and PhD from National Taiwan Univ., Univ. of Rhode Island, and Stanford Univ., all in Electrical Eng. He is an IEEE Fellow "for contributions to digital processing of speech".