Quality improvement of HMM-based synthesized speech based on decomposition of naturalness and intelligibility using non-negative matrix factorization

Dinh Anh Tuan (1410030)

School of Information Science,
Japan Advanced Institute of Science and Technology

February 10, 2016

Keywords: HMM-based synthesized speech, naturalness, intelligibility, decomposition, non-negative matrix factorization.

HMM-based synthesized voices are intelligible but not natural especially in limited data condition because of over-smoothing speech spectra. One solution for the problem is using voice conversion techniques to convert over-smoothed spectra to natural spectra. Although conventional conversion techniques transform speech spectra to natural ones to improve naturalness, they cause unexpected distortions on acceptable intelligibility of synthesized speech. The aim of this paper is to improve naturalness without violating acceptable intelligibility employing a novel asymmetric bilinear model (ABM) using non-negative matrix factorization (NMF) to separate the naturalness and intelligibility of synthesized speech. Subjective evaluations carried out on English data confirm that the achieved synthesis quality is higher than other methods in limited data condition and competitive in large data condition. Moreover, non-negativity constrain in NMF helps reveal the physical meanings of factored matrices as naturalness and intelligibility.