ACOUSTICS2008/1580
Statistical Conversion of Speech Parameter Trajectory for Mapping between Features of Different Modalities

Tomoki Toda
Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma, 630-0192 Nara, Japan

A state-of-the-art speech parameter conversion technique and its application to a mapping between features of different modalities are reviewed. Many statistical approaches to the parameter conversion have been studied particularly for voice conversion in speech synthesis research. A typical method conducts the parameter conversion frame by frame based on the minimum mean square error using a Gaussian mixture model of the joint probability density of input and output parameters [Y. Stylianou et al., IEEE Trans. SAP, Vol. 6, No. 2, pp. 131-142, (1998)]. Although this method is reasonably effective, the deterioration of the conversion accuracy is caused by essential problems of the frame-based conversion process. Recently a conversion method based on the maximum likelihood estimation of a parameter trajectory has been proposed [T. Toda et al., IEEE Trans. ASLP, Vol. 15, No. 8, pp. 2222-2235, (2007)]. This method realizes the appropriate converted parameter sequence by 1) using not only static but also dynamic feature statistics and 2) considering a global variance feature of the converted parameters. It has been reported that this method is effective in several applications such as a spectral determination from articulatory movements, an acoustic-to-articulatory inversion mapping, and a conversion of body-transmitted speech into air-transmitted speech.