

Deep Learning-Based Voice Conversion

Zhenhua Ling¹

概要 : I will introduce our recent work on applying deep learning techniques to voice conversion in this talk. Several methods have been proposed to improve different components in the pipeline of a statistical parametric voice conversion system, including deep neural networks with layer-wise generative training for acoustic modeling, deep autoencoders with binary distributed hidden units for feature representation, and WaveNet vocoder with limited training data for waveform reconstruction. Then, I will introduce our system designed for Voice Conversion Challenge 2018, which achieved the best performance under both parallel and non-parallel conditions in this evaluation. After this, I will present our recent progress on sequence-to-sequence acoustic modeling for voice conversion, which converts the acoustic features and durations of source utterances simultaneously using a unified acoustic model. Finally, some discussions on the future development of voice conversion techniques will be given.

経歴

Zhenhua Ling received the B.E. degree in electronic information engineering, the M.S. and Ph.D. degree in signal and information processing from the University of Science and Technology of China, Hefei, China, in 2002, 2005, and 2008, respectively. From October 2007 to March 2008, he was a Marie Curie Fellow with the Centre for Speech Technology Research, University of Edinburgh, Edinburgh, U.K. From July 2008 to February 2011, he was a joint Postdoctoral Researcher with the University of Science and Technology of China, Hefei, China, and iFLYTEK Co., Ltd., China. He is currently an Associate Professor with the University of Science and Technology of China. He also worked at the University of Washington, Seattle, WA, USA, as a Visiting Scholar from August 2012 to August 2013. His research interests include speech processing, speech synthesis, voice conversion, and natural language processing. He was the recipient of the IEEE Signal Processing Society Young Author Best Paper Award in 2010. He is now an Associate Editor for the IEEE/ACM transactions on audio, speech, and language processing.

¹ University of Science and Technology of China, China