Probabilistic Modeling of Speech

Professor Zhijian Ou
Visiting Scholar, Beckman Institute; University of Illinois

**ABSTRACT:** This talk will overview some of our recent work in SPMI (Speech Processing and Machine Intelligence) lab, Tsinghua University, under the thread of probabilistic modeling of speech. Most speech processing tasks (e.g. pitch estimation, speech recognition, source separation and so on) require a probabilistic model of speech. The more scientific the model is, the better we can do for speech processing.

1. In our ICASSP-2007 paper, we studied the Bayesian HMM modeling of speech, treating the concatenation of all the Gaussian means in a HMM as a random supervector. The idea of estimating utterance eigenvoices and performing (unsupervised) utterance adaptation is found to be useful.
2. In our ICASSP-2010 paper, we proposed variational nonparametric Bayesian Hidden Markov Model, and demonstrated its ability in discovering the structure of HMM's hidden state space for speech recognition.
3. In our ICASSP-2011 paper, we studied the NMF modeling of voice in song, and built a monaural voice and accompaniment separation system.
4. These pieces of previous work enlighten us to propose PAT (Probabilistic Acoustic Tube) model of speech in AISTATS-2012, ICASSP-2014, which becomes one of our main research topics. PAT is based on the fundamental physics of speech production, incorporating mixed excitation, glottal wave and phase modeling. We demonstrate the capability of PAT for a number of speech analysis/synthesis tasks, such as pitch tracking under both clean and additive noise conditions, speech synthesis, and phoneme clustering. One of the reviewers comments "to my knowledge the most complete attempt on developing a true generative model for speech."

**BIO:** Zhijian Ou received the B.S. degree with the highest honor in Electronic Engineering from Shanghai Jiao Tong University, Shanghai, China, in 1998, and the M.S. and Ph.D. degrees in Electronic Engineering from Tsinghua University, Beijing, China, in 2000 and 2003, respectively. Since 2003, he has been with the Department of Electronic Engineering, Tsinghua University, Beijing, China, and is now an Associate Professor. He has been working as the principle investigator for research projects from National Science Foundation China (NSFC), the High Technology Research and Development Program of China (863 project), and China Ministry of Information Industry, as well as for joint-research projects with Intel, Panasonic, IBM, and Toshiba. His current research interests include Speech Processing (speech recognition and understanding, source separation, speaker recognition, natural language processing), and Statistical Machine Intelligence (particularly with graphical models).

**MARCH 11 4 PM - 5 PM**

**4369 BECKMAN INSTITUTE**