Keynote Speech II

# Who Can Understand Your Speech Better — Deep Neural Network or Gaussian Mixture Model?

**Dr. Dong Yu,**
**Microsoft Research**

**Abstract:** Recently we have shown that the context-dependent deep neural network (DNN) hidden Markov model (CD-DNN-HMM) can do surprisingly well for large vocabulary speech recognition (LVSR) as demonstrated on several benchmark tasks. Since then, much work has been done to understand its potential and to further advance the state of the art. In this talk I will share some of these thoughts and introduce some of the recent progresses we have made.

In the talk, I will first briefly describe CD-DNN-HMM and bring some insights on why DNNs can do better than the shallow neural networks and Gaussian mixture models. My discussion will be based on the fact that DNN can be considered as a joint model of a complicated feature extractor and a log-linear model. I will then describe how some of the obstacles, such as training speed, decoding speed, sequence-level training, and adaptation, on adopting CD-DNN-HMMs can be removed thanks to recent advances. After that, I will show ways to further improve the DNN structures to achieve better recognition accuracy and to support new scenarios. I will conclude the talk by indicating that DNNs not only do better but also are simpler than GMMs.

**Bio:** Dr. Dong Yu joined Microsoft Corporation in 1998 and Microsoft Speech Research Group in 2002, where he is currently a senior researcher. He holds a PhD degree in computer science from University of Idaho, an MS degree in computer science from Indiana University at Bloomington, an MS degree in electrical engineering from Chinese Academy of Sciences, and a BS degree (with honors) in electrical engineering from Zhejiang University.  His recent work focuses on deep neural network and its applications to large vocabulary speech recognition. Dr. Dong Yu has published over 100 papers in speech processing and machine learning and is the inventor/co-inventor of around 50 granted/pending patents. He is currently serving as an associate editor of *IEEE transactions on audio, speech, and language processing* (2011-) and has served as an associate editor of *IEEE signal processing magazine* (2008-2011) and the lead guest editor of *IEEE Transactions on Audio, Speech, and Language Processing* special issue on deep learning for speech and language processing (2010-2011).