**2010-2011 ISCA Distinguished Lecture Trip-Two (Xi’An) Report**

**Lecturer:** Li Deng

**Local Host:** Professor Lei Xie, Northwest Polytech University, Xi’an, China   
  
**Travel and lecture information**

October 14, 2011: Arrived in Xi’an China and accommodated at Le Garden Hotel (hotel cost and all most local costs covered by the host).

Three ISCA Distinguished Lectures were delivered during the following week. Their titles and abstracts are as follows:

**Title:  Hierarchical Structures for Statistical Speech Modeling and Recognition**

Recently, there has been intense research in the use of deep architectures as hierarchical generative models for machine learning, information processing, and related applications in pattern synthesis and recognition. For speech researchers, it is natural to motivate the use of deep architectures from the perspectives of human speech production and perception. In this talk, I will analyze speech production and perception mechanisms and describe a series of hierarchical statistical models aimed to capture the relevant properties for the purpose of automatic speech recognition. I will argue that the currently dominant hidden Markov model (HMM) is a shallow architecture that can be much improved by exploiting speech production and perception hierarchies. Some early work on trajectory model, segmental model, switching dynamic system model, hidden dynamic model, hidden trajectory, and hybrid HMM and neural network model developed for speech recognition beyond the flat-structured HMM framework will be reviewed from the perspective of hierarchical modeling. More recent work that is motivated by a special type of deep architecture and has achieved striking success will be presented.

**Title:  Feature-Domain, Model-Domain, and Hybrid Approaches to Noise-Robust Speech Recognition**

Noise robustness has long been an active area of research that captures significant interest from speech recognition researchers and developers. In this lecture, we use the Bayesian framework as a common thread to connect, analyze, and categorize a number of popular approaches to noise robust speech recognition pursued in the recent past. The topics covered in this lecture include: 1) Bayesian decision rules with unreliable features and unreliable model parameters; 2) Principled ways of computing feature uncertainty using structured speech distortion models; 3) Use of phase factor in an advanced speech distortion model for feature compensation; 4) A novel perspective on model compensation as a special implementation of the general Bayesian predictive classification rule capitalizing on model parameter uncertainty; 5) Taxonomy of noise compensation techniques using two distinct axes: feature vs. model domain and structured vs. unstructured transformation; and 6) Noise adaptive training as a hybrid feature-model compensation framework and its various forms of extension.

**Title: A Tutorial on Deep Learning for Signal and Speech Information Processing**

Today, signal processing research has a significantly widened scope compared with just a few years ago, and machine learning has become an important technical area of our signal processing community. Since 2006, deep learning—a new area of machine learning research—has emerged, impacting a wide range of signal and information processing work within the traditional and the new, widened scopes. Various workshops, such as the 2011 ICML Workshop on Learning Architectures, Representations, and Optimization for Speech and Visual Information Processing, the 2009 ICML Workshop on Learning Feature Hierarchies, the 2008 NIPS Deep Learning Workshop, the 2009 NIPS Workshop on Deep Learning for Speech Recognition and Related Applications, as well as an upcoming special issue on Deep Learning for Speech and Language Processing in IEEE Transactions on Audio, Speech, and Language Processing (2011), have all been devoted exclusively to deep learning and its applications to various classical signal processing areas. We have also seen the government sponsor research on deep learning (e.g., the DARPA deep learning program). The main purpose of this tutorial is to review the basic mathematical  and machine learning framework, and to introduce the audience to the emerging technologies enabled by deep learning. I will also review the research work conducted in this area since the birth of deep learning in 2006, especially the work that is of direct relevance to signal processing. Future research directions will be discussed from my personal perspective. It is hopeful that this tutorial can attract interests from signal and speech processing researchers, students, and practitioners in this emerging area for advancing signal and information processing technology and applications in the future.

**Summary of outcomes:**

The impact of the lectures has been strong. Students and researchers had strong interactions with the lecturer. Importance of speech processing is emphasized and perceived well. ISCA is considered as the active promoter for serving developing countries well with the planned DL tours with the lecturers acting as an ambassador. The audience was very engaging and provided many valuable feedbacks on the need for different levels of education on speech processing.

The audience consists of students, educators, and industrial engineers. The all welcome continuing the DL program from ISCA.