ISCA DL Travel Report by Sadaoki Furui

From March 27, 2013, to March 31, I travelled to Thailand as an ISCA Distinguished Lecturer. I visited Bangkok, and gave a keynote talk at TSC-5 (5th Meeting of Thai Speech Community), held at Thammasat University, Bangkok. The cost for international flights was covered by our university, Tokyo Institute of Technology, and accommodation and food were covered by local hosts. My visit to Bangkok, Thailand, was very fruitful.

My lecture was well attended, and led to useful and inspiring discussions. I could not only encourage researchers and students to pursue research on automatic speech recognition, but also get new ideas for future research from various useful and positive feedback on my lecture. I could also visit Chulalongkorn University and NECTEC (National Electronics and Computer Technology Center) to have various technical discussions with faculty, researchers and students. I am planning to keep in touch with them, as it would be nice to set up some form of collaboration in the near future.

Details of the lecture are as follows:

Talk title: "Automatic speech recognition: trials, tribulations and triumphs"

Abstract:

   Although many important scientific advances have taken place in automatic speech recognition (ASR) research, we have also encountered a number of practical limitations which hinder a widespread deployment of applications and services. In most speech recognition tasks, human subjects produce one to two orders of magnitude fewer errors than machines. One of the most significant differences exists in that human subjects are far more flexible and adaptive than machines against various variations of speech, including individuality, speaking style, additive noise, and channel distortions. How to train and adapt statistical models for speech recognition using a limited amount of data is one of the most important research issues.

   What we know about human speech processing and the natural variation of speech is very limited. It is important to spend more effort to clarify especially the mechanism underlying speaker-to-speaker variability, and devise a method for simultaneously modeling multiple sources of variations based on statistical analysis using large-scale databases. Future systems need to have an efficient way of representing, storing, and retrieving various knowledge resources.

   Data-intensive science is rapidly emerging in scientific and computing research communities. The size of speech databases/corpora used in ASR research and development is typically 100 to 1,000 hours of utterances, which is too small considering the variety of sources of variations. We need to focus on solving various problems before efficiently constructing and utilizing huge speech databases, which will be essential to next-generation ASR systems.

Attendance: around 100 people from academia, industry and government

Host: Dr. Chai Wutiwiwatchai, NECTEC, Bangkok, chai.wutiwiwatchai@nectec.or.th.