ISCA DL Travel Report by Sadaoki Furui

From November 4, 2012, to November 18, I travelled to South America as an ISCA Distinguished Lecturer. I visited Ecuador, Argentina and Brazil, and gave three lectures/talks. The cost for international as well as domestic flights was covered by ISCA, and accommodation and food were covered by local hosts. My visit to South America was very fruitful.

All my lectures/talks were well attended, and led to useful and inspiring discussions. I could not only encourage faculty members 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 lectures/talks. I could also visit laboratories of host professors in Argentina and Brazil to have various technical discussions with faculty members and students. I was also given a tour of the department and campus, and introduced to the administration such as the department head and dean at each university. 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 three lectures/talks are as follows:

1) **Keynote talk at IEEE LATINCOM 2012 conference (4th IEEE Latin-American Conference on Communications 2012) held in the Convention Center of the Salesian Polithecnical University (UPS - Universidad Politécnica Salesiana) in Cuenca, Ecuador**

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 300 people from academia, industry and government

Host: Prof. Pedro Vizcaya Guarin, Pontificia Universidad Javeriana, Columbia, pvizcaya@javeriana.edu.co

Prof. Julio Cesar Viola, Department of Electronics and Circuits, University of Simon Bolivar, Venezuela, jcviola@usb.ve

2) Lecture at Buenos Aires Institute of Technology, Argentina

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

Abstract: same as above

Attendance: around 50 people, consisting of faculty members, researchers and students

Host: Prof. Roxana saint-Nom, Department of Electronics and Electric Engineering, Buenos Aires Institute of Technology, Argentina, saitnom@itba.edu.ar

3) Lecture at University of Sao Paulo, Brazil

Lecture title: "Data-intensive ASR based on machine learning"

Abstract:

Since speech is highly variable, even if we have a fairly large-scale database, we cannot avoid the data sparseness problem in constructing automatic speech recognition (ASR) systems. How to train and adapt statistical models using limited amounts of data is one of the most important research issues in ASR. This talk summarizes major techniques that have been proposed to solve the generalization problem in acoustic model training and adaptation, that is, how to achieve high recognition accuracy for new utterances. One of the common approaches is controlling the degree of freedom in model training and adaptation. The techniques can be classified by whether a priori knowledge of speech obtained from a speech database such as those recorded using many speakers is used or not. Another approach is maximizing "margins" between training samples and the decision boundaries. Many of these techniques have also been combined and extended to further improve performance.

Although many useful techniques have been developed, we still do not have a golden standard that can be applied to any kind of speech variation and any condition of the speech data available for training and adaptation. We need to focus on collecting rich and effective speech databases covering a wide range of variations, active learning for automatically selecting data for annotation, cheap, fast and good-enough transcription, and unsupervised, semi-supervised or lightly-supervised training/adaptation, based on advanced machine learning techniques. We also need to extend current efforts to understand more about human speech processing and the mechanism of natural speech variation.

Attendance: around 30 people, consisting of faculty members, researchers and students

Host: Prof. Vitor H. Nascimento, Department of Electronics Systems Engineering, University of Sao Paulo, Brazil, vitor@lps.usp.br