Distinguished Lecture Series

Selected topics from 40 years of my research on speech processing

**Monday, October 7th, 2013 10:00am
Auditorium 106 at New IIS Building**

Sadaoki Furui

**Tokyo Institute of Technology
Toyota Technological Institute at Chicago**

Abstract

This talk introduces major topics from my research on speech processing which I have conducted for more than 40 years, always bearing innovation in mind, at NTT (Nippon Telegraph and Telephone Corporation) Research Labs and Tokyo Institute of Technology (Tokyo Tech). I have conducted research mainly on automatic speech recognition (ASR), including automatic speaker recognition. ASR researchers have made significant progress since 1970 when I started my research, and now the techniques are successfully used in various application systems. I have been fortunate to have witnessed and participated in a period of significant technological progress.

In the 1970's, I conducted research on speaker recognition using statistical features and showed that speech variation over time is one of the most serious problems in speaker recognition. At the end of 1970's, I conducted research at Bell Labs and proposed a speaker recognition method using dynamic features of the speech spectrum. In the 1980's, I verified that dynamic features of the speech spectrum play an important role in phoneme perception by human beings, and applied the method using the combination of instantaneous and dynamic cepstrum, which I originally proposed for speaker recognition, to ASR. This method became a standard and is still widely used. For speaker recognition, we conducted pioneering research using GMMs. In the 1990's, we conducted various research on model adaptation for ASR, and proposed text-prompted speaker recognition. In the 2000's, I supervised a project on spontaneous speech recognition based on a large speech corpus, which significantly contributed to the progress of spontaneous speech recognition. In this project, we conducted pioneering research on automatic speech summarization. I also supervised a COE (Center of Excellence) project for large-scale knowledge resources at our university, and a national project for developing practical ASR techniques in cooperation with major Japanese companies. In the latter project, we successfully constructed a very high-performance WFST-based ASR decoder.

I believe a major direction of future ASR research is modeling of various speech variations based on advanced machine-learning techniques and efficient construction and utilization of large-scale speech corpora.

This talk also includes my recent activities at Tokyo Tech and Toyota Technological Institute at Chicago.

For more information: http://www.iis.sinica.edu.tw/