NATIONAL UNIVERSITY OF SINGAPORE

School of Computing

C S   S E M I N A R

Title: **1) Overview of Speech research at NTT - Tomohiro Nakatani, Senior Research Scientist 2) NTT system for the REVERB challenge - Marc Delcroix, Research Scientist 3) ASR error detection and recognition rate estimation using deep bidirectional recurrent neural networks - Atsunori Ogawa, Researcher 4) DOLPHIN - A statistical multichannel speech denoising approach based on spatial and spectral characteristics of sources - Tomohiro Nakatani, Senior Research Scientist**

Speaker: Tomohiro Nakatani
Senior Research Scientist
NTT Communication Science Laboratories
Kyoto, Japan

Marc Delcroix
Research Scientist
NTT Communication Science Laboratories
Kyoto, Japan

Atsunori Ogawa
Researcher
NTT Communication Science Laboratories
Kyoto, Japan

Date/Time: 27 April 2015, Monday, 10:00 AM to 12:00 PM

Venue: MR6, AS6-05-10

Chaired by: Dr Sim Khe Chai, Assistant Professor, School of Computing
(simkc@comp.nus.edu.sg)

Title: Overview of Speech research at NTT

Abstract:
In this talk we will give a brief overview of our main research activities. Our research projects include,microphone-array based speech enhancement, single channel speech enhancement, robust ASR, acoustic event detection etc.

Biodata:
Tomohiro Nakatani received the B.E., M.E., and Ph.D. degrees from Kyoto University, Kyoto, Japan, in 1989, 1991, and 2002, respectively. He is a Senior Research Scientist

(Supervisor) of NTT Communication Science Laboratories, Kyoto, Japan. Since joining NTT in 1991, he has been investigating speech enhancement technologies for developing intelligent human-machine interfaces. Since 2008, he has been a Visiting Assistant Professor in the Department of Media Science, Nagoya University. He is currently an associate member of IEEE SP Society Audio and Acoustics Technical Committee.

Title: NTT system for the REVERB challenge

Abstract:
Reverberation is one of the factors limiting performance of distant automatic speech recognition systems. The REVERB challenge was recently organized to evaluate the progress in the field of reverberant speech enhancement and recognition. In this talk we introduce the system that we proposed for the REVERB challenge, which achieved high recognition performance even in severe reverberant conditions. Our system consists of a speech enhancement pre-processor, followed by a DNN-based recognizer. We describe the different components of our system and detail the contribution of each component on our final results.

Biodata:
Marc Delcroix is a research scientist at the media information laboratory of NTT Communication Science Laboratories, Kyoto, Japan. He received the M.Eng. degree from the Free University of Brussels, Belgium, and the Ecole Centrale Paris, France, in 2003 and the Ph.D.degree from the Graduate School of Information Science and Technology, Hokkaido University, Japan, in 2007. His research interests include speech enhancement, dereverberation and robust speech recognition.

Title: ASR error detection and recognition rate estimation using deep bidirectional recurrent neural networks

Abstract:
Recurrent neural networks (RNNs) have recently been applied as the classifiers for sequential problems. In this talk, deep bidirectional RNNs (DBRNNs) are applied for the first time to error detection in automatic speech recognition (ASR), which is a sequential labeling problem. We investigate three types of ASR error detection tasks, i.e. confidence estimation, out-of-vocabulary word detection and error type classification. We also estimate recognition rates from the error type classification results. Experimental results show that the DBRNNs greatly outperform conditional random fields (CRFs), especially for the detection of infrequent error labels. The DBRNNs also slightly outperform the CRFs in recognition rate estimation.

Biodata:
Atsunori Ogawa received the B.E. and M.E. degrees in information engineering, and the Ph.D. degree in information science from Nagoya University, Aichi, in 1996, 1998, and 2008, respectively. He is currently a researcher at NTT Communication Science Laboratories, Kyoto, Japan, and engaged in the research on automatic speech recognition. He is a member of the IEEE, Institute of Electronics, Information, and Communications Engineers (IEICE), Information Processing Society of Japan (IPSJ), and the Acoustical

Society of Japan (ASJ). He received the ASJ Best Poster Presentation Award in 2003 and 2006, respectively.

Title: DOLPHIN - A statistical multichannel speech denoising approach based on spatial and spectral characteristics of sources

Abstract:
In this talk, a statistical multichannel speech denoising approach,referred to as DOLPHIN, is presented. The approach optimally utilizes the spatial and spectral characteristics of speech and background noise, and can distinguish them precisely even in highly non-stationary noise environments. As an example, we focus on a challenge benchmark for Machine Listening in Multisource Environments, referred to as CHiME Challenge, and show how DOLPHIN can work effectively for this benchmark and improves the performance of the ASR backend. Some extensions for more general tasks and future work are also mentioned at the end of the talk.

Biodata:
Tomohiro Nakatani received the B.E., M.E., and Ph.D. degrees from Kyoto University, Kyoto, Japan, in 1989, 1991, and 2002, respectively. He is a Senior Research Scientist (Supervisor) of NTT Communication Science Laboratories, Kyoto, Japan. Since joining NTT in 1991, he has been investigating speech enhancement technologies for developing intelligent human-machine interfaces. Since 2008, he has been a Visiting Assistant Professor in the Department of Media Science, Nagoya University. He is currently an associate member of IEEE SP Society Audio and Acoustics Technical Committee.