



A Deep Learning Approach to Speech Enhancement and Source Separation

by

Professor Chin-Hui Lee

School of Electrical and Computer Engineering
Georgia Institute of Technology
USA



Date: 9 November 2015 (Monday)

Time: 10 a.m. – 11:30 a.m.

Venue: Room 222, 2/F, Ho Sin Hang Engineering Building, CUHK

Abstract

In contrast to conventional theory-based signal processing, we formulate a novel deep learning framework as finding a regression function between an observed signal and a corresponding set of desired targets. Monte Carlo and big data techniques are often required to generate a large collection of input and output pairs in order to learn the often-complicated structure of the mapping functions. In this talk, we present recent advances in our lab on speech enhancement and source separation. Other related issues in speech recognition, speech synthesis, voice conversion, voice activity detection, speech dereverberation and bandwidth expansion will also be highlighted. Leveraging upon the recently emerging big data paradigm, many classical signal processing problems can now be cast in machine learning frameworks with deep neural networks (DNNs) capable of joint training of both continuous speech parameters, such as ideal binary mask in enhancement, log power spectrum in separation, and categorical information, such as voicing, phonation position and phone identity in synthesis. We believe a combination of deep learning and big data offers vast opportunities for R&D professionals to explore many ideas that are not easily accomplished in the past and pioneer new grounds that are still wide open in the future.

Biography of the Speaker

Chin-Hui Lee is a professor at School of Electrical and Computer Engineering, Georgia Institute of Technology. Dr. Lee received the B.S. degree in Electrical Engineering from National Taiwan University, Taipei, in 1973, the M.S. degree in Engineering and Applied Science from Yale University, New Haven, in 1977, and the Ph.D. degree in Electrical Engineering with a minor in Statistics from University of Washington, Seattle, in 1981.

Dr. Lee started his professional career at Verbex Corporation, Bedford, MA, and was involved in research on connected word recognition. In 1984, he became affiliated with Digital Sound Corporation, Santa Barbara, where he engaged in research and product development in speech coding, speech synthesis, speech recognition and signal processing for the development of the DSC-2000 Voice Server. Between 1986 and 2001, he was with Bell Laboratories, Murray Hill, New Jersey, where he became a Distinguished Member of Technical Staff and Director of the Dialogue Systems Research Department. His research interests include multimedia communication, multimedia signal and information processing, speech and speaker recognition, speech and language modeling, spoken dialogue processing, adaptive and discriminative learning, biometric authentication, and information retrieval. From August 2001 to August 2002 he was a visiting professor at School of Computing, The National University of Singapore. In September 2002, he joined the Faculty Georgia Institute of Technology.

Dr. Lee is a Fellow of IEEE and ISCA. He has published over 400 papers and 30 patents, and accumulated over 25,000 citations with an h-index of 62 on Google Scholar. He received the SPS Senior Award in 1994 and the SPS Best Paper Award in 1997 and 1999. In 1997, he was awarded the prestigious Bell Labs President's Gold Award for his contributions to the Lucent Speech Processing Solutions product. Dr. Lee often gives seminal lectures to a wide international audience. In 2000, he was named one of the six Distinguished Lecturers by the IEEE Signal Processing Society. He was also named one of the two ISCA's inaugural Distinguished Lecturers in 2007-2008. He won the SPS's 2006 Technical Achievement Award for "Exceptional Contributions to the Field of Automatic Speech Recognition". In ICASSP2012 he gave a plenary talk on the future of automatic speech recognition. In the same year he was awarded the ISCA Medal in scientific achievement for "pioneering and seminal contributions to the principles and practice of automatic speech and speaker recognition, including fundamental innovations in adaptive learning, discriminative training and utterance verification".