ViVoLab Speech Technologies Group

Instituto Universitario de Investigación en Ingeniería de Aragón

Universidad Zaragoza

IRIS Towards Natural Interaction and Communication

MARIE CURIE ACTIONS

INDUSTRY-ACADEMIA PARTNERSHIPS AND PATHWAYS (IAPP) - MARIE CURIE ACTIONS

Speech Technologies Seminar

Dear Colleagues, It is our great pleasure to **invite** you to attend to the speech technologies seminar with

Prof. Chin-Hui Lee and **Prof. Richard Stern**.

Date and time: *Wednesday 28th May 2014, 11 am*
Location: *Salón de actos Edificio I+D+i, 1ª planta,*
C/Mariano Esquillor s/n,
Campus Rio Ebro, Zaragoza
Contact person: *Prof. Eduardo Lleida*

The seminar is free and open to the public.

11:00	**Prof. Chin-Hui Lee** School of Electrical and Computer Engineering, Georgia Institute of Technology http://users.ece.gatech.edu/chl/ **A Deep Neural Network Approach to Speech Enhancement**
12:15	Coffee break
12:30	**Prof. Richard Stern** Department of Electrical and Computer Engineering and Language Technologies Institute, Carnegie Mellon University http://users.ece.cmu.edu/~rms/ **Applying Physiologically-Motivated Models of Auditory Processing to Automatic Speech Recognition: Promise and Progress**

A Deep Neural Network Approach to Speech Enhancement

In contrast to the conventional minimum mean square error (MMSE) based noise reduction techniques, we formulate speech enhancement as finding a mapping function between noisy and clean speech signals. In order to be able to handle a wide range of additive noises in real-world situations, a large training set, encompassing many possible combinations of speech and noise types, is first designed. Next a deep neural network (DNN) architecture is employed as a nonlinear regression function to ensure a powerful modeling capability. Several techniques have also been adopted to improve the DNN-based speech enhancement system, including global variance equalization to alleviate the over-smoothing problem of the regression model, and dropout and noise-aware training strategies to further improve the generalization capability of DNNs to unseen noise conditions. Experimental results demonstrate that the proposed framework can achieve significant improvements in both objective and subjective measures over the MMSE based techniques. It is also interesting to observe that the proposed DNN approach can well suppress the highly non-stationary noise, which is tough to handle in general. Furthermore, the resulting DNN model, trained with artificial synthesized data, is also effective in dealing with noisy speech data recorded in real-world scenarios without generating the annoying musical artifact commonly observed in conventional enhancement methods.

Applying Physiologically-Motivated Models of Auditory Processing to Automatic Speech Recognition: Promise and Progress

For many years the human auditory system has been an inspiration for developers of automatic speech recognition systems because of its ability to interpret speech accurately in a wide variety of difficult acoustical environments. This talk will discuss the application of physiologically-motivated and psychophysically-motivated approaches to signal processing that facilitates robust automatic speech recognition. The talk will begin by reviewing selected aspects of auditory processing that are believed to be especially relevant to speech perception, and that had been components of signal processing schemes that were proposed in the 1980s. We will review and discuss the motivation for, and the structure of, classical and contemporary computational models of auditory processing that have been applied to speech recognition, and we will evaluate and compare their impact on improving speech recognition accuracy. We will discuss some of the general observations and results that have been obtained during the renaissance of activity in auditory-based features over the past 15 years. Finally, we will identify certain attributes of auditory processing that we believe to be generally helpful, and share insights that we have gleaned about auditory processing from recent work at Carnegie Mellon.