**ISCA DL Travel Report by Hynek Hermansky**

From November 24, 2013, to December 2nd, I travelled to India as an ISCA Distinguished Lecturer. I visited Delhi and Hyderabad, and gave three lectures, including a keynote talk at Oriental COCOSDA 2013 (host Prof. Shyam Agrawal), held at KITT College of Engineering in Guargaon, close to Delhi. The other talks were a three hour tutorial on “Speech Recognition” at Indian Institute for Information Technology (IIIT) Hyderabad (host Prof. B. Yegnanarayana), and one hour lecture on “Time in Recognition of Speech” (host Prof. Murthy) at Indian Institute of Technology (IIT) Hyderabad. The cost for international flights was covered by my own funds, and internal flight in India, accommodation and food were covered by local hosts.

I had a great and ample opportunity to interact with students and senior researchers in India, trying to generate interest in processing of speech as a scientific discipline, and promoting ISCA activities. It seems very likely that I will be further interact with the newly established Digital Signal Processing Center at IIIT Hyderabad as a member of the Advisory Board on the Center.

The abstract of my keynote speech at COCOSDA 2013 is below

**Artificial Neural Networks: Deep, Long and Wide**

Up to recently, most current speech processing techniques build on a single stream of feature vectors. Each feature vector represents a relative short (10-20 ms) snapshots of speech signal. Recognizing that information is carried in changes of the speech signal, the short-term features are typically appended by dynamic features that indicate local spectral dynamics. The talk reviews historical reasons for use of the short-term features and points to basic problems with such an approach. The history of spectral representations of speech is briefly discussed. Some of the history of gradual infusion of the modulation spectrum concept into Automatic recognition of speech (ASR) is mentioned. Arguments for application of multiple processing streams in ASR are brought up. Finally, spectral dynamics-based phoneme posterior features that are organized in multiple processing streams are described and latest trends in their applications in machine recognition of speech are discussed, emphasizing applications dealing with limited amounts of training data in target languages

The abstract of my 3 hour tutorial at IIIT Hyderabad is below

Information Extraction from Speech: History and Some Current Techniques

The tutorial covers history of processing of speech, and some results from human speech perception, leading to currently dominant techniques for acoustic processing of speech signals in machine recognition of speech. The emphasis is on techniques from our research group, which are motivated by some basic properties of human speech processing.

The talk was video-recorded at the recording is available

The 1 hour talk at IIT Hyderabad was an abbreviate version on by IIIT tutorial