

Invited Talks in WiSSAP 2019

| Speakers | Topic | Description | Duration |
|--|---|---|----------|
| Ratnavalli Ellajosyula, Senior Consultant Neurologist & Specialist in Cognitive Neurology, Manipal Hospital & Annasawmy Mudaliar General Hospital | Bilingualism and dementia: The vulnerable second language | The study investigated language profiles in multilingual patients with Alzheimer's disease (AD), behavioural variant Frontotemporal dementia (bvFTD) and semantic dementia (SD) to explore if there is differential decline in secondary languages and the role of language proficiency. Striking loss of naming and comprehension in second language (L2) was seen in SD suggesting a common semantic system for different languages. L2 also appears to deteriorate faster compared to L1 in AD, suggesting increased vulnerability of L2 to brain degeneration. Language proficiency, age of acquisition and usage of language are possibly equally important factors, which require further study. | 90 mins |
| V Ramasubramanian, IIIT Bangalore | Deep Learning: The biology and mathematics beneath | <p>Deep learning has established itself strongly as a machine-learning paradigm both in theory and practice across a wide range of data domains and applications, especially for almost all aspects of machine processing of speech. The central theme of this talk is on how directions of research in auditory neuroscience in human-hearing – blending neurophysiological results and computational realizations – would pave the way for ever more successful computational equivalents of human speech processing and machine-perception algorithms and systems. Towards this, this talk will touch on two classes of deep learning – the Deep Neural Networks (DNNs) and the Convolution Neural Networks (CNNs or ConvNets) – and examine this central theme of how these architectures have a strong biological and mathematical basis.</p> <p>In the case of DNNs, this will specifically point to the 'mathematical basis' in the form of Cybenko's Universal Approximation Theorem for functional approximation, and the 'biological basis' built on single-neuron, perceptron and MLP models to classify linearly/non-linearly separable classes, even while also acknowledging the oft cited criticism of the backpropagation learning of DNNs not having a pleasing biological realism.</p> <p>In the case of CNNs (or ConvNets) this talk will point to its strong biological and mathematical basis, especially in machine vision domain, where the convolutional and pooling layers are directly inspired by the classic notions of simple cells and complex cells (e.g. orientation specific receptive fields) in visual neuroscience (Nobel prize winning work of Hubel and Wiesel, 1962), the corresponding computational equivalent of David Marr's formulations of edge detection and primal sketches in early visual processing and the current understanding of what ConvNets learn in their layers.</p> | 90 mins |
| Deepu Vijayasanan, NIT Surathkal | Speaker Representation -- from i-Vectors to end to end systems | The deep learning revolution has introduced many robust representations other than the conventional i-vectors. This talk would cover some of the state of the art methods in deriving speaker embeddings from speech utterances and how they are used in speaker verification tasks | 90 mins |

| | | | |
|---|--|---|-------------------|
| Jithendra Vepa, Observe.AI | Deep learning in Automatic Speech Recognition | This talk will start with overview of automatic speech recognition (ASR) systems and provide details of current ASR production systems particularly used in personal voice assistants. I will discuss about the latest research, mainly End-to-end ASR and briefly touch upon the challenges being addressed by the community to commercialize this technology. Finally, I will describe the techniques used for wake-word recognition and end-point detection, which are critical components for hands-free voice interface. | 90 mins |
| Chandra S Seelamantula, IISc, Bangalore | Learning to do sparse coding | In this talk, we shall explore the link between iterative algorithms deep networks for solving the problem of sparse coding. | 90 mins |
| Rajakrishnan Rajkumar, IISER Bhopal | Eye movements in the perception of Indian languages | The talk will provide an introduction to the neural and physiological mechanisms underlying the process of reading written text and illustrate how eye movements reflect linguistic processing. Subsequently, I will summarize existing research on Hindi sentence comprehension with particular reference to psycholinguistic works from recent years as well as readability research in Bangla. Finally, I will discuss how predictors of reading difficulty integrated into machine learning models have contributed to engineering applications in the field of Natural Language Processing. | 90 mins |
| Lori L. Holt, Carnegie Mellon University | Understanding how humans interpret the complexity of spoken language; Part I: Cracking the Speech Code with Learning, Part II: Stability and Plasticity in Human Speech Perception | <p>Experience deeply shapes how human listeners perceive spoken language. We learn long-term phonetic representations and words that respect the sound structure of our native language and, yet, we maintain enough flexibility to make sense of experience with nonnative accents or speech from imperfect computer synthesis. There are rich behavioral-science literatures that speak to the many ways that experience shapes speech perception. Yet, for the most part, contemporary neurobiological models of spoken language are oriented toward characterization of the system in a stable state. We are just beginning to understand the learning mechanisms involved in supporting successful human speech communication. I will describe how experience shapes speech perception at different time scales - from the influence of a single precursor sound, to distributions of sounds across seconds, to statistical regularities in acoustics experienced across multiple training sessions.</p> <p>In Part I, I will describe current thinking in how human listeners discover functional units in speech like phonemes and words and how this learning fundamentally shapes perception. In Part II, I will describe more dynamic aspects of speech comprehension that depend on very rapid adaptation and learning at short timescales. In general, this research demonstrates that human speech recognition is a flexible, adaptive, experience-dependent skill that draws upon perceptual, cognitive, motor and linguistic systems. I will argue that human speech communication has much to offer machine listening and speech recognition and that - reciprocally - next-generation approaches to human speech processing will benefit a great deal from closer connection to machine systems.</p> | 90 mins + 90 mins |

| | | | |
|---|---|--|-------------------|
| Barbara Shinn-Cunningham, Carnegie Mellon University | Role of attention mechanisms in listening | Understanding speech in natural environments depends not just on decoding the speech signal, but on extracting the speech signal from a mixture of sounds. In order to achieve this, the listener must be able to 1) parse the scene, determining what sound energy belongs to the speech signal and what energy is from a competing source (perform auditory scene analysis), and 2) filter out the competing source energy and focus on the speech. Together, these processes allow a listener to focus attention on the speech and analyze its content in detail. In Part I of my presentation, I will illustrate these issues, including what acoustic features support auditory scene analysis and what features allow a listener to focus attention. In Part II, I will describe the different brain networks that control auditory attention, and how we measure the effects of attention on neural processing. | 90 mins + 90 mins |
| Tijmen Tieleman, minds.ai | Getting creative with autoencoders | This presentation describes an autoencoder with a hard-coded domain-specific decoder, for the purpose of achieving powerful unsupervised learning. The unsupervised learning can be visualized quite interestingly, and can be used as a transfer learning source for supervised learning. | 90 mins |