

Audio visual Integration in Humans and Machines

Sriram Ganapathy (EE) & SP Arun (CNS)

Despite tremendous advances in computing, vision and speech are fundamentally unsolved computational problems. They are challenging for the brain too: large parts of the brain are involved in vision and speech processing, but we do not yet understand the nature of these representations. In addition, the human processing of audio-visual signals involving both visual and auditory processing is an even bigger puzzle. A fundamental, unanswered question regarding these representations is whether they are compositional: that is, whether the response to a complex object can be broken down into its constituent parts. This has been difficult to answer because natural visual and auditory stimuli cannot be easily and uniquely broken down into their parts. The long-term goal of this research is to understand how audio and visual signals are integrated in the brain. A parallel question in the artificial intelligence world is whether advanced deep learning models for audio-visual signals obey rules of compositionality. Insights from this research have important applications in making machines interact with humans and in advancing neuroscience.

This research will involve recording and analyzing brain signals using behaviour, EEG or MRI in humans as well as building computational deep learning models for audio-visual scenes or movie clips. The fundamental question is the understanding of the complex interactions between different modalities in the perception of such data for humans and machines. The ideal candidate should have an appetite for both neuroscience and machine learning and should have a strong background in computational and mathematical concepts like probability and linear algebra.

About the faculty

- *Sriram Ganapathy* is an Assistant Professor in the Department of Electrical Engineering, IISc and leads the activities of the learning and extraction of acoustic patterns (LEAP) lab, IISc. Previously, he was a research staff member at the IBM Watson Research Center for four years. He received his PhD from the Center of Language and Speech Processing, Johns Hopkins University. His research interests include signal processing and deep learning applied to speech recognition, speaker recognition and auditory neuroscience. He is a member of the ISCA, a senior member of the IEEE and subject editor for Elsevier Speech Communications. For more details, please visit the [LEAP webpage](#).
- *SP Arun* is an Associate Professor at the Centre for Neuroscience, IISc. He received his B.Tech from IIT Bombay, and MS & PhD from Johns Hopkins University, all in Electrical Engineering. He completed his postdoctoral training at Carnegie Mellon University and then joined the Centre for Neuroscience at IISc. His research interests lie broadly in understanding how the brain transforms sensations into perception. For more details visit the homepage of his research group, the [Vision Lab @ IISc](#).