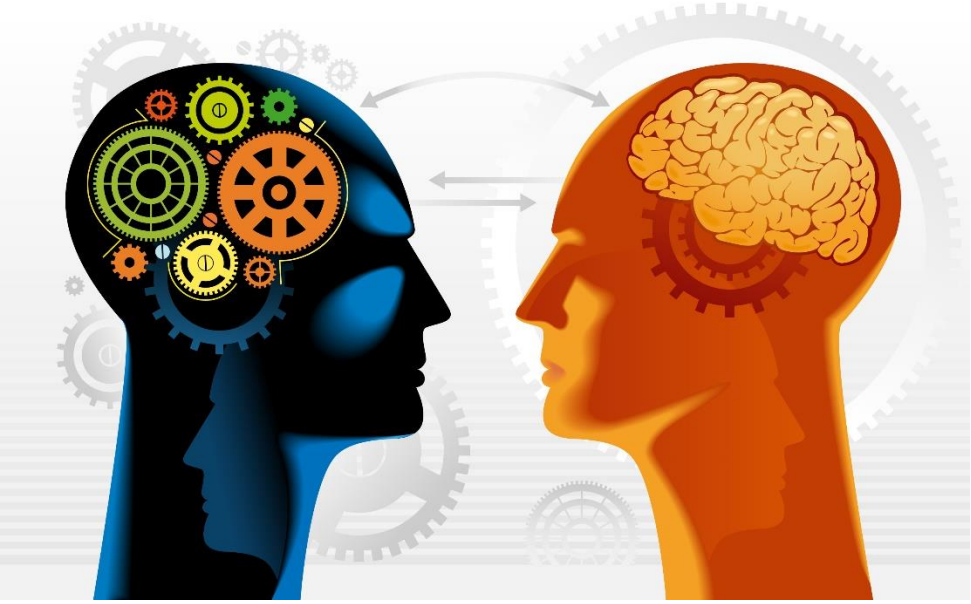
Learning audio-visual representations in deep networks and humans

Sriram Ganapathy (EE) & SP Arun (CNS)



Despite tremendous advances in computing, understanding human visual and auditory representations is a fundamentally unsolved problem. 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 multi-modal signals (audio-visual) involving both visual and auditory areas 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. The long-term goal of this research is to understand how audio and visual signals are integrated in the brain.

In the area of artificial intelligence, machine learning, powered with deep neural networks has made remarkable progress in the last decade, to achieve human level performance in many visual and audio tasks. Some of the key questions in the artificial this world are i) what representations are learned by the complex models and are there parallels to the representations learned in the brain, ii) how the decisions made by these systems can be explained, and iii) whether advanced deep learning models for audio-visual signals obey rules of compositionality. Insights from this research have important applications in building next generation of machine learning systems and in advancing neuroscience. This research will involve recording and analyzing brain signals using behavior, EEG or MRI in humans as well as building computational deep and convolutional neural models for audio-visual data.

Eligibility

The ideal candidate should have an appetite for both neuroscience and machine learning and should have a strong background in computational and mathematical concepts.

More Information

Webpage of Sriram Ganapathy | Webpage of SP Arun