

Rahul Sharma

3740 McClintock Ave, EEB 400, Los Angeles, CA, 90007
Email: rahul.sharma@usc.edu
Mobile No.: +12132456486

ACADEMIC DETAILS

Degree	Institute	Year	GPA
PhD in Electrical and Computer Engineering	USC	2017-Present	3.8/4
Integrated BTech – MTech in Electrical Engineering	IIT Kanpur	2012-17	8.8/10

FIELDS OF INTEREST

Primarily interested in **multimodal signal processing**, more inclined towards the **visual signal**, to understand human actions and behavior in multimedia content. Furthermore, I am keenly interested in **semi-supervised systems** and the notion of **weaker-than-full supervision**.

SCHOLASTIC ACHIEVEMENTS

- **Viterbi Graduate Student Fellowship (2017)**: Viterbi School of Engineering, USC.
- **Cadence Gold Medal (2017)**: Awarded by IIT Kanpur, for best master's thesis work across all departments.
- **GATE Fellowship (2016-2017)**: Awarded by HRD India, towards a stipend during the master's program.
- **Merit Cum Means Scholarship (2013 - 2016)**: Awarded by IIT Kanpur to support the tuition at the institute.

PUBLICATIONS

- **R. Sharma, S. Narayanan**, "[Crossmodal Learning for Audio-Visual Speech Event Localization](#)": arXiv preprint
- **A. Hebbar, R. Sharma**, "[Vocal Tract Articulatory Contour Detection in Real-Time Magnetic Real-Time MRI Using Spatio-Temporal Context](#)": Proceedings of International Conference on Acoustics, Speech and Signal Processing, May 2020
- **R. Sharma, S. Narayanan**, "[Towards Visual Voice Activity Detection for Unconstrained Videos](#)": Proceedings of International Conference on Image Processing, September 2019
- **R. Sharma, T. Guha, G. Sharma**, "[Multichannel Attention Network for Analyzing Visual Behavior in Public Speaking](#)": Proceedings of Winter Conference on Applications of Computer Vision, February 2018
- **R. Sharma, T. Guha**, "[A Trajectory Clustering Approach to Crowd Flow Segmentation in Videos](#)", Proceedings of International Conference on Image Processing, September 2016

CURRENT RESEARCH

- **Active Speaker Detection in Unconstrained Videos**: Working towards detecting an active speaker in each frame of the videos from Hollywood movies. We formulate it with a primary objective of visual voice activity detection (VAD) and a lateral objective of active speaker detection. We hypothesize that, to decide for VAD, the network should consider the active speakers as most salient. We exploit the interpretability of the employed 3-D convolutional network to localize the active speakers in the frames.
- **Tracking Vocal Tract Air-Tissue Contours in rt-MRI videos**: A step towards understanding the human vocal system using rt-MRI videos of midsagittal plane of human face. We develop a system to automatically detect the contours of the articulators involved in human speech synthesis. We train a CNN-LSTM architecture which utilizes the spatial and temporal information to reliably track the deforming articulators.
- **Video Understanding for Assisting Autism Diagnosis**: An effort to automatically understand the key characteristics of ASD kids' behavior as with respect to TD kids using the videos of home and clinical sessions. The system involves classifying the humans present in the scene into child and interlocutor. Further it involves computing the features representing the child-interlocutor dynamics.

MASTER'S THESIS

Towards Multimodal Assessment of Speaker Performance in Public Speaking (2017).

We propose a computational framework for quantifying speaker performance in the context of public speaking. For this purpose, we created a database consisting of more than two thousand Technology, Entertainment, Design (TED) conference videos along with associated metadata (number of likes/dislikes, views, comments) from YouTube. We do not consider the content of the talk, analyze the speech and the visual content to capture the verbal and non-verbal behavior of the speaker. We have established baselines which can predict the performance rating with correlation coefficient 0.68.

Published in WACV' 2018