Chaitanya Ahuja

5719 Gates and Hillman Center – Carnegie Mellon University

cahuja@andrew.cmu.edu • www.chahuja.com • chahuja

Education

Carnegie Mellon University — **Pittsburgh**
Ph.D. in Language Technologies — *Aug 2015 – Present*
Advisor: Louis-Philippe Morency

Indian Institute of Technology, Kanpur — **Kanpur**
B.Tech. in Electrical Engineering, 9.5/10 — *Aug 2011 – May 2015*
Minor in Artificial Intelligence

Research Areas

Pose generation, Multimodal Grounding, Multimodal Representation Learning, Speech Processing and Synthesis, Structured Prediction, Spatial Audio

Research Experience

Carnegie Mellon University, Prof. Louis-Philippe Morency — *August 2015 – Present*
Scalable Grounding of Human Body Motion on Spoken Language

- **Grounding Gestures:** Learnt a common embedding space for language, acoustics and human body pose for for the purposes of co-speech gesture generation.
- **Gesture Style Transfer and Control:** Transferred idiosyncratic gesture style of one speaker to the other in a many-to-many style transfer setup.
- **Learning from Limited Data:** *(Upcoming)* Learning to generate co-speech gestures for new speakers with significantly lesser supervision.

IIT Kanpur, Prof. Rajesh Hegde — *Aug 2014 – May 2015*
Final Year Project: *Source Separation using a Complex Matrix Factorization approach for Joint Modeling of Magnitude and Phase* [arXiv]

- Proposed a new algorithm to **jointly model magnitude and phase** while matrix factorization
- Reduced the Complex Matrix Factorization (CMF) problem to a simple Non-Negative Matrix Factorization (NMF) problem by simple transformations
- Justified the algorithm's effectiveness by comparing against state of the art source-separation methods
- Demonstrated that accurate phase reconstruction resolves unwanted artifacts in the reconstructed speech signal

IIT Kanpur, Prof. Vinay Namboodiri — *Aug 2014 – May 2015*
Final Year Project: *Visual Summarization of foreground object motion using boundary initialization of object tracking* [tech. report]

- Proposed an online system for creating **human-centric image summaries** of **surveillance videos** which is based on Kernel-based tracking for automated live synthesis of video synopsis of surveillance videos
- Initialized foreground objects based on locally varying blob-detection algorithm
- Clustered tracks based on time and space to prevent occlusion in the summary
- Video Summary was synthesized by placing objects, equally spaced in time, on the background

Selected Honors and Awards

- **Summer Undergraduate Research Grant for Excellence** (SURGE) 2013, IIT Kanpur
- One of the top 7 projects (out of 70) **in SURGE 2013**
- **Academic Excellence Award** for distinctive performance in terms 2011-12, 2012-13.
- **All India Rank 231 - Top 0.05%** (amongst 4,75,000 students) in IIT-JEE 2011.
- **All India Rank 124 - Top 0.05%** (amongst 10,00,000 students) in AIEEE 2011.

Publications

Published

[P1] C. Ahuja, D. W. Lee, Y. I. Nakano, and L.-P. Morency, "Style transfer for co-speech gesture animation: A multi-speaker conditional mixture approach," *ECCV*, 2020.

[P2] C. Ahuja, S. Ma, L.-P. Morency, and Y. Sheikh, "To react or not to react: End-to-end visual pose forecasting for personalized avatar during dyadic conversations," *ICMI*, 2019.

[P3] C. Ahuja and L.-P. Morency, "Language2pose: Natural language grounded pose forecasting," *3DV*, 2019. [Online]. Available: https://arxiv.org/pdf/1907.01108.pdf.

[P4] T. Baltrusaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *TPAMI*, 2018. [Online]. Available: https://arxiv.org/abs/1705.09406.

[P5] C. Ahuja and L.-P. Morency, "Lattice recurrent unit: Improving convergence and statistical efficiency for sequence modeling," *AAAI*, 2018. [Online]. Available: https://arxiv.org/abs/1710.02254.

[P6] C. Ahuja and R. M. Hegde, "Fast modelling of pinna spectral notches from hrtfs using linear prediction residual cepstrum," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2014, pp. 4458–4462. [Online]. Available: http://chahuja.com/files/icassp_chahuja_paper.pdf.

[P7] A. Sohni, C. Ahuja, and R. M. Hegde, "Extraction of pinna spectral notches in the median plane of a virtual spherical microphone array," in *4th Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, IEEE, 2014, pp. 142–146. [Online]. Available: http://chahuja.com/files/hscma_chahuja_paper.pdf.

Book Chapters

[B0] T. Baltrusaitis, C. Ahuja, and L.-P. Morency, "Challenges and applications in multimodal machine learning," in *The Handbook of Multimodal-Multisensor Interfaces: Signal Processing, Architectures, and Detection of Emotion and Cognition-Volume 2*, 2018, pp. 17–48.

Tech. Reports

[A0] C. Ahuja, K. Nathwani, and R. M. Hegde, "A complex matrix factorization approach to joint modeling of magnitude and phase for source separation," *arXiv preprint arXiv:1411.6741*, 2014. [Online]. Available: https://arxiv.org/abs/1411.6741.

Internships

Facebook Reality Labs, Shugao Ma *Summer 2018*

Human Communication Dynamics

- In a dyadic conversation setting, designed a model to generate upper body animations conditioned on the avatar's speech, pose history and interlocutor's speech and pose history.
- Used an attention based model to focus on interpersonal and intrapersonal dynamics as and when indicated by the stimuli to the model.
- Demonstrated the model's effectiveness in generating accurate and natural looking pose sequences via various objective and subjective metrics of evaluation.

Cornell University, Prof. Tsuhan Chen *Summer 2014*

Prediction of Adjectives for given Nouns using Probability distribution of adjective-noun pairs and adjective-adjective similarity [tech. report]

- Designed a system to **predict adjectives** for a given noun based on an existing set of tags, which increased the vocabulary of the tags while maintaining the sanctity of the noun-adjective pair
- Incorporated a Sentence Corpus (British-National-Corpus) to improve the compatibility of adjective with respect to nouns based on a probability measure
- **Removed redundant data** from the sentence corpus using a hash table which increased accuracy as compared to the baseline

IIT Kanpur, Prof. Rajesh Hegde *Summer 2013*

On-Line modeling of the Pinna for Computation of HRTF's in Rendering 3D Audio

- Explored relations between structure of a ear and Head Related Transfer Functions (HRTFs)
- Worked towards mimicking a ear with digital filters to synthesize **Spatial Audio**
- Developed methods to verify ear contours generated by spectral notches of HRTFs, hence mapping HRTFs to the anthropometry of the ear.

Teaching and Professional Activities

- **Reviewer** ICMI *2020*
- **Reviewer** SIGGRAPH *2020*
- **Reviewer** Association for Computational Linguistics (ACL) *2020*
- **Reviewer** Empirical Methods in Natural Language Processing (EMNLP) *2019*
- **Reviewer** Association for Computational Linguistics (ACL) *2019*
- **Teaching Assistant** Structured Prediction for Language and Other Discrete Data (CMU 11-763) *Spring 2018*
- **Teaching Assistant** Advanced Multimodal Machine Learning (CMU 11-777) *Spring 2017*
- **Reviewer** International Conference on Learning Representations (ICLR) *2017*
- **Reviewer** NIPS Workshop on Multimodal Machine Learning *2016*

Selected Course Projects

Deep RL and control *Jan 2017 – May 2017*

- *Segmentation Models for NLP tasks with RL* [tech. report]

Segmenting sentences into useful phrases for tasks like Machine Translation and Summarization

Statistical Machine Learning *Jan 2017 – May 2017*

- *Topological Data Analysis* [tech. report] [presentation]

Analysing confidence intervals in cluster trees to facilitate pruning of low-confidence branches (or leaves)

Multimodal Machine Learning *Aug 2015 – May 2016*

- *Video Captioning* [tech. report]

Generating descriptive captions for movie video segments.

Skills

- Languages: Bash, C, CSS, HTML, LaTeX, Make, Python
- Frameworks: Numpy, Pandas, Pytorch, Scipy, Scikitlearn, Tensorflow, Theano
- OS: Linux, OSX

Graduate Coursework

- Deep Reinforcement Learning (CMU 10-703): R. Salakhutdinov, K. Fragkiadaki *Spring 2017*
- Statistical Machine Learning (CMU 10-702): L. Wasserman, R. Tibshirani *Spring 2017*
- Deep Learning (CMU 10-707): R. Salakhutdinov *Fall 2016*
- Intermediate Statistics (CMU 10-705): L. Wasserman *Fall 2016*
- Advanced Multimodal Machine Learning (CMU 11-777): L.-P. Morency *Spring 2016*
- Machine Learning (CMU 10-701): T. Mitchell *Spring 2016*
- Human Communication and Multimodal ML (CMU 11-776): L.-P. Morency *Fall 2015*
- Algorithms for NLP (CMU 10-702): C. Dyer *Fall 2015*