

Chaitanya Ahuja

5719 Gates and Hillman Center – Carnegie Mellon University

✉ cahuja@andrew.cmu.edu • chahuja.com • [chahuja](https://github.com/chahuja) • [chahuja](https://twitter.com/chahuja)

Education

Ph.D. in Language Technologies

School of Computer Science, Carnegie Mellon University, 3.97/4.0

Advisor: Prof. Louis-Philippe Morency

Pittsburgh, PA

Aug 2015 – Dec 2021 (expected)

B.Tech. in Electrical Engineering

Indian Institute of Technology Kanpur, 9.5/10

Minor: Artificial Intelligence

Kanpur, India

Aug 2011 – May 2015

Research Interests

- **Generative Modeling:** addressing multimodal grounding, translation, alignment, fusion, and low-resource learning.
- **Virtual Presence:** multimodal virtual agents with natural verbal (i.e., speech, language, spatial audio) and nonverbal behaviours (i.e., hand gestures, facial expressions, body pose) for remote interaction, education and collaboration
- **Social Intelligence:** multimodal embodied agents that can perceive and interact with humans in real or virtual environments

PhD Thesis Research

Communication beyond words: Grounding Visual Body Motion with Language

Carnegie Mellon University, Prof. Louis-Philippe Morency

- **Multimodal Grounding:** Constructed a common embedding space for language, acoustics and human body pose for the purposes of co-speech gesture generation.
- **Gesture Style Transfer and Control:** Proposed a many-to-many gesture style transfer set-up and designed efficient approaches to transfer the idiosyncratic gesture styles of one speaker to another speaker.
- **Low-Resource Generative Models:** Designed approaches to generate co-speech gestures for new speakers with significantly lesser supervision.

Research Positions

- Graduate Researcher, Carnegie Mellon University
Advisor: Prof. Louis-Philippe Morency
Pittsburgh, PA
2015 – Present
- Research Intern, Facebook Reality Labs
Advisor: Shugao Ma
Pittsburgh, PA
Summer 2018
- Research Intern, Cornell University
Advisor: Prof. Tsuhan Chen
Ithaca, PA
Summer 2014
- Undergraduate Researcher, Indian Institute of Technology
Advisor: Prof. Rajesh Hegde
Kanpur, India
2013 – 2015
- Undergraduate Researcher, Indian Institute of Technology
Advisor: Prof. Vinay Namboodiri
Kanpur, India
2014 – 2015

Honors and Awards

- CMU Graduate Research Fellowship, 2015 – 2020
- Honorable Mention at the LTI Student Research Symposium, 2019
- Cornell-India summer program at Cornell University, 2014
- Viterbi-India summer program at the University of Southern California (declined), 2014

- o Summer Undergraduate Research Grant for Excellence ([SURGE](#)) 2013, IIT Kanpur
- o One of the top 7 projects (out of 70) in [SURGE 2013](#)
- o Academic Excellence Awards for distinctive performance, 2011 – 2013, IIT Kanpur
- o All India Rank 231 - Top 0.05% (amongst 4,75,000 students) in IIT-JEE 2011.
- o All India Rank 124 - Top 0.05% (amongst 10,00,000 students) in AIEEE 2011.

Publications

Updated list on [Google Scholar](#)

Pre-prints/Under review.....

1. Low-Resource Adaptation of Spatio-Temporal Crossmodal Generative Models
Chaitanya Ahuja, Dong Won Lee, Louis-Philippe Morency
Under Review
2. Crossmodal clustered contrastive learning: Grounding of spoken language to gestures
Dong Won Lee, **Chaitanya Ahuja**, Louis-Philippe Morency
Under Review

Refereed conferences/journals.....

1. [No gestures left behind: Learning relationships between spoken language and freeform gestures](#)
Chaitanya Ahuja, Dong Won Lee, Ryo Ishii, Louis-Philippe Morency
Findings of Empirical Methods in Natural Language Processing (Findings of EMNLP, 2020)
Presented at Natural Language Beyond Text Workshop @EMNLP 2020
[\[code\]](#)
2. [Impact of personality on nonverbal behavior generation](#)
Ryo Ishii, **Chaitanya Ahuja**, Yukiko I. Nakano, Louis-Philippe Morency
ACM International Conference on Intelligent Virtual Agents (IVA, 2020)
3. [Style transfer for co-speech gesture animation: A multi-speaker conditional mixture approach](#)
Chaitanya Ahuja, Dong Won Lee, Yukiko I. Nakano, Louis-Philippe Morency
European Conference on Computer Vision (ECCV, 2020)
[\[code\]](#) [\[demo\]](#) Media: [TechXplore](#)
4. [To react or not to react: End-to-end visual pose forecasting for personalized avatar during dyadic conversations](#)
Chaitanya Ahuja, Shugao Ma, Louis-Philippe Morency, Yaser Sheikh
ACM International Conference on Multimodal Interaction (ICMI, 2019) ([oral](#))
5. [Coalescing Narrative and Dialogue for Grounded Pose Forecasting](#)
Chaitanya Ahuja
Doctoral Consortium, ACM International Conference on Multimodal Interaction (ICMI 2019)
6. [Language2pose: Natural language grounded pose forecasting](#)
Chaitanya Ahuja, Louis-Philippe Morency
International Conference on 3D Vision (3DV, 2019)
[\[code\]](#) [\[webpage\]](#) Media: [Scientific American](#), [Synced](#), [VentureBeat](#)
7. [A complex matrix factorization approach to joint modeling of magnitude and phase for source separation](#)
Chaitanya Ahuja, Karan Nathwani, Rajesh M. Hegde
IEEE International Symposium on Signal Processing and Information Technology (ISSPIT, 2019)
8. [Multimodal machine learning: A survey and taxonomy](#)
Tadas Baltrušaitis, **Chaitanya Ahuja**, Louis-Philippe Morency
Transactions on Pattern Analysis and Machine Intelligence (TPAMI, 2018)
9. [Lattice recurrent unit: Improving convergence and statistical efficiency for sequence modeling](#)
Chaitanya Ahuja, Louis-Philippe Morency
AAAI Conference on Artificial Intelligence (AAAI, 2018)
[\[code\]](#) [\[webpage\]](#)
10. [Fast modelling of pinna spectral notches from HRTFs using linear prediction residual cepstrum](#)
Chaitanya Ahuja, Rajesh M. Hegde

Book Chapters

1. [Challenges and applications in multimodal machine learning](#)
Tadas Baltrušaitis, **Chaitanya Ahuja**, Louis-Philippe Morency
The Handbook of Multimodal-Multisensor Interfaces: Signal Processing, Architectures, and Detection of Emotion and Cognition-Volume 2, 2018, pp. 17–48

Refereed Workshops

1. [Extraction of pinna spectral notches in the median plane of a virtual spherical microphone array](#)
Ankit Sohni, **Chaitanya Ahuja**, Rajesh M. Hegde
Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA, 2014)

Organized Workshops

1. [First Workshop on Crossmodal Social Animation](#)
International Conference on Computer Vision (ICCV, 2021) Workshop Proceedings

Technical Reports

1. [Training Segmentation Models for Extractive and Generative NLP Tasks with Reinforcement Learning](#)
Akash Bharadwaj*, **Chaitanya Ahuja***
Course Project, Deep RL and control at CMU
2. [Topological Data Analysis](#)
Bhuwan Dhingra*, **Chaitanya Ahuja***
Course Project, Statistical Machine Learning at CMU [\[slides\]](#)
3. [Video Captioning](#)
Salvador Medina*, **Chaitanya Ahuja***
Course Project, Advanced Multimodal Machine Learning at CMU
4. [Visual Summarization of foreground object motion using boundary initialization of object tracking](#)
Chaitanya Ahuja*, Pratik Somani*
B.Tech Final Year Project at IIT Kanpur

Talks

- o **Communication Beyond Words: Grounding Visual Body Motion with Spoken Language**
KTH Stockholm, Online *April 2021*
- o **Learning Relationships between Spoken Language and Freeform Gestures**
EMNLP 2020 Workshop on NLP Beyond Text, Online *November 2020*
- o **Natural Language Grounded Pose Forecasting**
LTI Student Research Symposium, Pittsburgh PA *August 2019*
- o **End-to-End Visual Pose Forecasting for Personalized Avatar during Dyadic Conversations**
ACM International Conference on Multimodal Interaction, Suzhou, China *October 2019*

Student Mentorship

- o Dong Won Lee (CMU BS → CMU MS in Machine Learning): Self-supervised generative models.
- o Shradha Sehgal (IIIT Hyderabad B.Tech.): Evaluation of generative models.
- o Arvin Wu (CMU BS): Social intelligence benchmarking.
- o Nikitha Murikinati (CMU BS): Study of relationships between co-speech gestures and prosody.
- o Sharath Rao (CMU MS → PlayStation): Back-channel prediction in dyadic conversations.
- o Qingtao Hu (CMU MS → Amazon): Unsupervised disentanglement of style and content in images.
- o Anirudha Rayasam (CMU MS → Google): Language grounded pose forecasting.

Teaching

- o Head TA: 11763 [Structured Prediction for language and discrete data](#) by Taylor Berg-Kirkpatrick and Bhiksha Raj, CMU
3 recitations on Viterbi Decoding [\[slides\]](#), ILP and Dependency Parsing [\[slides\]](#) and Neural CRFs [\[slides\]](#) *Spring 2018*
- o Head TA: 11-777 Multimodal Machine Learning by Louis-Philippe Morency, CMU *Spring 2017*

Resources

- o **PATS Dataset**: Designed and constructed a large benchmark to study the complex multimodal relationships between Body Poses, Audio, Transcripts, and individual gesture Styles

Professional Activities and Service

- o Co-organizer: ICCV 2021 First Workshop on Crossmodal Social Animation
- o Co-organizer: Multimodal Machine Learning Reading Group, CMU, Spring 2020
- o Conference Program Committee: NeurIPS, SIGGRAPH, ICLR, ACL, EMNLP, ACM Multimedia, ICMI
- o Workshop Program Committee: NeurIPS workshop on Multimodal Machine Learning, ACL Workshop on Multimodal Language, NAACL-HLT Student Research Workshop, ICMI GENE Workshop
- o Grant Reviewer: Army Research Office (ARO)
- o CMU Graduate Applicant Support Program Volunteer: 2020
- o CMU AI Undergraduate Research Mentor: 2020, 2021
- o CMU Graduate Student Association Representative for Language Technologies Institute: 2017

Skills

- o Languages: English (fluent), Hindi (native), Spanish (Limited Working)
- o Programming Languages: Python, C, MATLAB, CSS, HTML, \LaTeX
- o Frameworks: Numpy, Pandas, PyTorch, Scikit-Learn, Scipy, Tensorflow, Theano

Relevant Graduate Coursework

- o Structured Prediction for Language and Other Discrete Data (CMU 10-763): T. Berg-Kirkpatrick, B. Raj *Spring 2018*
- o Deep Reinforcement Learning (CMU 10-703): R. Salakhutdinov, K. Fragkiadaki *Spring 2017*
- o Statistical Machine Learning (CMU 10-702): L. Wasserman, R. Tibshirani *Spring 2017*
- o Deep Learning (CMU 10-707): R. Salakhutdinov *Fall 2016*
- o Intermediate Statistics (CMU 10-705): L. Wasserman *Fall 2016*
- o Advanced Multimodal Machine Learning (CMU 11-777): L.-P. Morency *Spring 2016*
- o Machine Learning (PhD) (CMU 10-701): T. Mitchell *Spring 2016*
- o Human Communication and Multimodal ML (CMU 11-776): L.-P. Morency *Fall 2015*
- o Algorithms for NLP (CMU 10-702): C. Dyer *Fall 2015*

References

Available on request