HUMAN COMMUNICATION AND MULTIMODAL MACHINE LEARNING

Description | **Schedule** | **Bibliography** | **Grades**

Instructors: Prof. Louis-Philippe Morency, morency@cs.cmu.edu, 412-268-5508
Office: GHC-5411
Time: Tuesdays and Thursday, 4:30pm-6pm
Classroom: GHC-4201

Recommended preparation: This is a graduate course primarily for students in LTI, HCII and Robotics; others, for example (undergraduate) students of CS or professional masters, by prior permission of the instructor. Students should have proper academic background in probability, statistic and linear algebra. Previous experience in machine learning is suggested but not obligatory. Programming knowledge in Matlab and/or Python is recommended.

Introduction and Purposes

Human face-to-face communication is a little like a dance, in that participants continuously adjust their behaviors based on verbal and nonverbal displays and signals. Human interpersonal behaviors have long been studied in linguistic, communication, sociology and psychology. The recent advances in machine learning, pattern recognition and signal processing enabled a new generation of computational tools to analyze, recognize and predict human communication behaviors during social interactions. This new research direction have broad applicability, including the improvement of human behavior recognition, the synthesis of natural animations for robots and virtual humans, the development of intelligent tutoring systems, and the diagnoses of social disorders (e.g., autism spectrum disorder).

The objectives of this course are:

(1) To give a general overview of human communicative behaviors (verbal, vocal and visual) and show a parallel with computer science subfields (natural language processing, speech processing and computer vision);
(2) To understand the multimodal challenge of human communication (e.g. speech and gesture synchrony) and learn about multimodal computational modeling;
(3) To understand the social aspect of human communication and its implication on statistical and probabilistic modeling;
(4) To learn about recent advances in multimodal machine learning and pattern recognition to analyze, recognize and predict human communicative behaviors;
(5) To give students practical experience in computational study of human social communication through a course project.

Course format

Lectures are twice a week on Tuesdays and Thursdays at 4pm. Tuesday lectures and the first hour of Thursday lecture will be given given by Prof. Morency or one of the guest lecturers. The 30 minutes of Thursday lectures will be discussions about the assigned research papers. Two or more students will be assigned to lead each discussion.

Course Material

Required:

- Reading material will be based on published technical papers available via the ACM/IEEE/Springer digital libraries or freely available online. All CMU students have automatic access to these digital archives.
- Matlab software, using CMU license (for practical exercises) or Python

Optional:

- *Multimodal Processing and Interaction,* Gros, Potamianos and Maragos, SpringerLink, 2008, DOI: 10.1007/978-0-387-76316-3 (freely available on SpringerLink for CMU students)
- *Nonverbal Communication in Human Interaction (7th edition), Mark Knapp and Judith Hall, Wadsworth, 2010*
- *Speech and Language Processing (2nd edition), Daniel Jurafsky and James Martin, Pearson, 2008*
- *Machine Learning for Audio, Image and Video Analysis: Theory and Applications*, Francesco Camastra and Alessandro Vinciarelli, Springer, 2008, DOI: 10.1007/978-1-84800-007-0 (freely available on SpringerLink for CMU students)

Course Topics and Readings

** Topics and reading assignments may change based on student interests and time restrictions. **

Classes	Lectures	Tentative reading list	Discussion leaders
Week 1 8/31	**Introduction and communication models** • Human communication dynamics • Signals and communicative messages • Contextual factors (task, environment, culture) • Communication models (Brunswick's model) • Syllabus and course project • Datasets and applications		
Week 2 9/7	**Visual messages** • Gesture, gaze, posture and proxemics • Facial expressions and coding systems • Object and activity recognition • Image and video representation	**Introduction and communication models** • •	• Karen Chen •
Week 3 9/14 *** Project pre-proposals due*	**Vocal messages** • Phonetics and phonology • Prosody and voice quality • Vocal expressions • Audio representation and basic feature extraction	**Visual messages** • •	• Behnaz Nojavan

Week 4 9/21	**Psychological and affective states** • Affect, moods and emotions • Emotion models (appraisal, PAD, categorical) • Attitude and personality traits (Big Five model) • Mental health and clinical diagnosis (DSM-5) • Education and learning metrics	**Vocal messages** • •	• Vivian Robinson
Week 5 9/28	**Statistical Evaluation and Analysis** • Hypothesis building and testing • Independent and dependent variables • Ground truth and coder agreement (or not!) • Statistical analysis (ANOVA, t-test, effect-size) • Crowdsourcing and wisdom of crowd	**Psychological and affective states** • •	• Leah Nicolich-Henlin • Tanmay Sinha
Week 6 10/5 *** Project proposals due 10/11*	**Behavior representation and analysis** • Feature representation and statistics • Dimensionality reduction and clustering • Canonical correlation analysis • Feature selection and mutual information • Dynamic time warping	**Statistical evaluation and analysis** • •	• Hongliang Yu • Arjun Bhardwaj
Week 7 10/12	**Multimodal machine learning: basic concepts** • Review of basic probability concepts • Classification regression and evaluation • Hyper-parameters and cross-validation • Bayes rule and Naïve Bayes classifier • Static discriminative models (decision tree, SVM)	**Behavior representation and analysis** • •	• Chaitanya Ahuji • Hiroaki Hayashi

Week 8 10/19	**Verbal messages** • Language models and N-grams • Boundaries, fillers and disfluencies • Syntax and part-of-speech tagging • Dependency tree parsing • Sphinx, hTK and syntax parsers	**Multimodal machine learning: basic concepts** • •	• Sz-Rung Shiang
Week 9 10/26 *** Mid-term reports due 10/30*	**Conversational messages** • Discourse analysis • Turn-taking and backchannel • Semantics and pragmatics • Speech and dialogue acts	**Verbal messages** • •	• Ran Zhao • Yoichi Matsuyama
Week 10 11/2	**Multimodal fusion and generative models** • Multimodal fusion • Audio-visual recognition • Hidden Markov Models • Multi-streams, coupled, factorial and asynchronous HMMs	**Conversational messages** • •	• Aaksha Maghawat • Junwei Liang

Week 11 11/9	**Multi-view conditional models** • Maximum entropy models • Conditional random fields • Latent-dynamic CRF • Multi-view HCRF model	**Multimodal fusion and generative models** • •	• Salvador Medina • Shikun Zhang
Week 12 11/16	**Multimodal deep learning** • Fuzzy machine learning • Neural networks • Deep belief networks • Auto-encoders	**Multi-view conditional models** • •	• Kazuya Kawakami • Pallavi Baljekar
Week 13 11/23	**Recurrent neural networks** • Long short term memory model • Recurrent neural networks • Thanksgiving weekend	**Multimodal deep learning** • •	• Sandeep Subramanian • Volkan Cirik
Week 14 11/30	**Final project presentations**		
Week 15 12/7 *** Final projects due 12/13*	**Final project presentations**		

Bibliography

Reading list from 2014 edition

Introduction and communication models

1. Morency, L.-P., Modeling Human Communication Dynamics, IEEEE Signal Processing Magazine, September 2010
2. A. Vinciarelli, M. Pantic and H. Bourlard, Social Signal Processing: Survey of an Emerging Domain, in Image and Vision Computing Journal, vol. 27, no. 12, pp. 1743-1759, December 2009
3. Krauss, R.M. (2002). The psychology of verbal communication. In, N. Smelser & P. Baltes (eds.), *International Encyclopedia of the Social and Behavioral Sciences.* London: Elsevier.
4. (optional) Pentland, Honest Signals, Chapter 1

Vocal messages

5. Schuller et al., (2011), Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge, Volume 53, Issues 9–10, 2011, Pages 1062–1087
6. Bachorowski et al. (2001), The acoustic features of human laughter, *Journal of the Acoustic Society of America*, 110 (3), pp. 1581-1597
7. (Optional) Ladefoged (2004), A course in phonetics
8. (Optional) Jurafsky and Martin (2008), Speech and Language Processing, Chapter 7, Sections 7.1-7.4

Visual messages

9. Krämer, N. C. (2008). Nonverbal Communication. In J. Blascovich & C. Hartel (eds.), Human behavior in military contexts (pp. 150-188). Washington: The National Academies Press. [Blackboard]
10. Fernando de la Torre and Jeffrey F. Cohn, Facial Expression Analysis, *Visual Analysis of Humans*, 2011, 377-409
11. (optional) Adam Kendon, An Agenda for Gesture Studies, This article appeared in Volume 7 (3) of the Semiotic Review of Books.
12. (optional) Michael Argyle and Janet Dean, Eye-contact, distance and Affiliation, Sociometry, Vol. 28, No. 3, pp. 289-304, 1965

Study Design, Evaluation and Analysis

13. Gale M. Lucas, Jonathan Gratch, Aisha King, Louis-Philippe Morency, It's only a computer: Virtual humans increase willingness to disclose, Computers in Human Behavior, Volume 37, August 2014, Pages 94-100.
14. H Wainer - How to display data badly - The American Statistician, 1984
15. (optional) Leroy (2011), Designing User Studies in Informatics, Springer.
16. (optional) http://www.sparknotes.com/psychology/psych101/researchmethods/

Multimodal Machine Learning: Basic Concepts

17. J.-I. Biel and D. Gatica-Perez, The YouTube Lens: Crowdsourced Personality Impressions and Audiovisual Analysis of Vlogs ,IEEE Trans. on Multimedia, Vol. 15, No. 1, pp. 41-55, Jan. 2013
18. Tom Fawcett, An introduction to ROC analysis, *Pattern Recognition Letters,* Volume 27, Issue 8, June 2006, Pages 861–874

19. (optional) Langley and Kibler (1991), The Experimental Study of Machine Learning, Unpublished paper.

Behavior analysis and unsupervised learning

20. Xuran Zhao, Nicholas Evans, and Jean-Luc Dugelay. 2012. CO-LDA: A Semi-supervised Approach to Audio-Visual Person Recognition. In Proceedings of the 2012 IEEE International Conference on Multimedia and Expo (ICME '12). IEEE Computer Society, Washington, DC, USA, 356-361. DOI=10.1109/ICME.2012.14 http://dx.doi.org/10.1109/ICME.2012.14
21. F. Zhou, F. De la Torre and J. F. Cohn (2010), Unsupervised Discovery of Facial Events, IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

Affective messages and personality traits

22. de Melo, C. M., Carnevale, P. J., Read, S. J., & Gratch, J. (2013, September 30). ReadingPeople's Minds From Emotion Expressions in Interdependent Decision Making. Journal of Personality and Social Psychology. Advance online publication. doi: 10.1037/a0034251
23. Jonathan Gratch, Lin Cheng, Stacy Marsella and Jill Boberg, Felt emotion and social context determine the intensity of smiles in a competitive video game, Face and Gesture 2013
24. (optional) Gratch & Marsella, 2005 Lessons from Emotion Psychology for the Design of Lifelike Characters
25. (optional) Mr Barrick, Mk Mount (1991) The Big Five Personality Dimensions And Job Performance: A Meta-Analysis - Personnel Psychology

Verbal messages

26. Pang et al. EMNLP 2002: Thumbs up? Sentiment classification using machine learning techniques, http://www.aclweb.org/anthology/W02-1011
27. Andreas Stolcke , Noah Coccaro , Rebecca Bates , Paul Taylor , Carol Van Ess-Dykema , Klaus Ries , Elizabeth Shriberg , Daniel Jurafsky , Rachel Martin , Marie Meteer, Dialogue act modeling for automatic tagging and recognition of conversational speech, Computational Linguistics, v.26 n.3, p.339-373, September 2000
28. (optional) Jurafsky and Martin (2008), Speech and Language Processing, Sections 4.1-4.4, 5.1-5.3 and 12.1-12.2 [Blackboard]
29. (optional) Soo-Min Kim and Eduard Hovy (2004) Determining the Sentiment of Opinions, Proceedings of the COLING conference, Geneva
30. (optional) Yang Liu, Elizabeth Shriberg, Andreas Stolcke, Dustin Hillard, Mari Ostendorf, Barbara Peskin, and Mary Harper. 2004. The ICSI-SRI-UW Metadata Extraction System, ICSLP 2004

Conversational messages

31. Duncan (1974) Some Signals and Rules for Taking Speaking Turns in Conversations
32. Bohus, D., Horvitz, E., (2010) - Computational Models for Multiparty Turn-Taking, Microsoft Technical Report MSR-TR-2010-115
33. (optional) Jurafsky and Martin (2008), Speech and Language Processing, Sections 17.1-17.4 and 21.1-21.4 [Blackboard]
34. (optional) Clark and Brennan (1991) Grounding in Communication, [Blackboard]

Multimodal behavior recognition (1/3)

35. C. Christoudias, K. Saenko, L.-P. Morency, and T. Darrell (2006) Co-Adaptation of Audio-Visual Speech and Gesture Classifiers, International Conference on Multimodal Interactions (ICMI 2006)

36. I. McCowan, D. Gatica-Perez, S. Bengio, G. Lathoud, M. Barnard, M., D. Zhang, "Automatic analysis of multimodal group actions in meetings", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 27, pp. 305–317, 2005
37. (optional) Gros, Potamianos and Maragos (2008) Multimodal Processing and Interaction, SpringerLink, Chapter 1 [SpringerLink or Blackboard]
38. (optional) Pradeep K. Atrey, M. Anwar Hossain, Abdulmotaleb El Saddik and Mohan S. Kankanhalli, Multimodal fusion for multimedia analysis: a survey, Multimedia Systems, Volume 16, Number 6 (2010), 345-379 (optional) A. Nefian, L. Liang, X. Pi, X. Liu and K. Murphy, (2002) Dynamic Bayesian networks for audio-visual speech recognition, EURASIP Journal on Applied Signal Processing, Volume 2002, Issue 1

Multimodal behavior recognition (2/3)

39. Konstantinos Bousmalis, Louis–Philippe Morency and Maja Pantic, Modeling Hidden Dynamics of Multimodal Cues for Spontaneous Agreement and Disagreement Recognition, Face and Gestures 2011
40. (optional) Rana El Kaliouby and Peter Robinson (2005) Real-Time Inference of Complex Mental States from Facial Expressions and Head Gestures, Proceedings of the workshop on Real-Time Vision for Human-Computer Interaction
41. (optional) Y. Tong, J. Chen & Q. Ji (2010), "A Unified Probabilistic Framework for Spontaneous Facial Action Modeling and Understanding", *IEEE Transactions on Pattern Analysis and Machine Intelligence.*
42. (optional) L.-P. Morency, A. Quattoni and Trevor Darrell (2007), Latent-Dynamic Discriminative Models for Continuous Gesture Recognition, Proceedings IEEE Conference on Computer Vision and Pattern Recognition, June 2007

Multimodal behavior recognition (3/3)

43.
44. Scherer, S., Glodek, M., Schwenker, F., Campbell, N. and Palm, G. (2012), Spotting Laughter in naturalistic multiparty conversations: a comparison of automatic online and offline approaches using audiovisual data*, ACM Transactions on Interactive Intelligent Systems: Special Issue on Affective Interaction in Natural Environments*, 2 (1), pp. 4:1-4:31.
45. (optional) McNeill, D. (1985). "So you think gestures are nonverbal?" In *Psychological Review* 92:350-371.
46. (optional) Scherer, S., Kane, J., Gobl, C. and Schwenker, F. (2012), Investigating Fuzzy-Input Fuzzy-Output Support Vector Machines for Robust Voice Quality Classification*, Computer Speech and Language*, in print.
47. (optional) Nalini Ambady and Robert Rosenthal (1992) Thin-slices of Expressive Behavior as Predictor of Interpersonal Consequences: A Meta-Analysis, Psychological Bulletin, Vol. 111, No. 2, 256-274
48. (optional) P. Verlinde and G. Chollet (1999), Comparing decision fusion paradigms using k-NN based classifiers, decision trees and logistic regression in a multi-modal identity verification application, Proceedings of the International Conference on Audio and Video-Based Biometric Person Authentication

Grades

- **Grading breakdown**
  - Participation 10%
  - Reading assignments 25%
  - Leading class discussion 15%
  - Course project:
    - Pre-proposal 5%
    - Proposal 5%
    - Mid-term report 15%
    - Final report 15%
    - Presentation 15%

- **Participation**
  - Students are expected to actively participate during group discussions and in class.
- **Reading assignments**
  - The reading assignment for each class will consist of 1-3 research papers (posted online one week before the class). These papers are specially selected to complement the lectures and show state-of-the-are research.
  - Friday before each class, 1-3 questions will be posted online. Students are required to post at least twice on the forum each week. These posts can be direct answers to the original questions or follow-up to other student posts.
  - Students must post their answers by 4pm on the Tuesday of each week. The answers will often be part of the group discussion.
- **Group discussions**
  - Each student will be leading the group discussion once during the semester. A signup sheet will be available during the first class.
  - Students can lead the discussion individually or pair with another student.
  - Since all students are expected to read the research papers, the discussion should bring something new and interactive to the class. This includes: example datasets, simple implementation of the algorithms, demo, new challenging questions and applications.

- **Course project:**
  - The goal of this course is to analyze human communicative behaviors in social settings using state-of-the-art statistical and probabilistic models. The course project is specifically design to give students practical experience in computational study of human social communication.
  - Students can perform the project individually or in teams of two. The mid-term and final report will need to outline the tasks of each student. Project reports for teams of 2 will be expected to include a deeper analysis than individual projects.
  - Students are free to select any topic for their course project as long as it is directly aligned with the course (multimodal computation and social interaction). The pre-proposal required early in the semester is designed to give early feedback to the students.
  - **Mid-term report:** The mid-term report will present a qualitative analysis of the selected dataset and communicative behaviors. The report should include correct transcription and annotations of the verbal, vocal and visual behaviors. Using standard statistical tools and qualitative observations, the students should highlight the challenges with this dataset (and communicative behaviors) and suggest an approach to solve them.
  - **Final report and presentation:** Using the same dataset as the mid-term report, the final report will include a quantitative analysis of the human communicative behaviors. The final report should be phrase as a research paper describing either a comparative study of different statistical and probabilistic approaches or a new technique for behavior modeling.