MULTIMODAL PROBABILISTIC LEARNING OF HUMAN COMMUNICATION

Description | **Schedule** | **Bibliography** | **Requirements** | **Notices**

Time: Wednesdays 3:30-6:20pm
Classroom: GFS 118
Instructors:
Professor Louis-Philippe Morency, morency@ict.usc.edu, 310-448-5323
Dr. Stefan Scherer, scherer@ict.usc.edu
Office: ICT 333

Recommended preparation: CSCI 542 or CSCI 567 or CSCI 573 or equivalent. Students should have proper academic background in probability, statistic and linear algebra. Previous experience in machine learning is suggested but not obligatory. This course is not a replacement for the Machine Learning course (CSCI 567).

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 (language, vocal and nonverbal) 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 signal processing;

(3) To understand the social aspect of human communication and its implication on statistical and probabilistic modeling;

(4) To learn about recent advances in 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

Each class will be three hours including two short pauses. The first two hours will consist of lectures given by Prof. Morency or one of the guest lecturers. The last hour will be a discussion about the assigned research papers. Two 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 USC students have automatic access to these digital archives.
- Matlab software, using USC license (for practical exercises)

Optional:

- *Multimodal Processing and Interaction,* Gros, Potamianos and Maragos, SpringerLink, 2008, DOI: 10.1007/978-0-387-76316-3 (freely available on SpringerLink for USC 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 USC students)

Course Topics and Readings

** Topics and readings may change based on student interest **

Classes	Lectures (2:00pm-3:50pm)	Readings for discussion sessions (4:00pm-4:50pm)	Discussion leaders
Week 1	**Introduction and communication models** • A multi-modal, multi-party, multi-label		

	dynamic problem • Human communication dynamics • Applications and domains • Communication models • Mid-term and final projects • Datasets and sensing tools		
Week 2	**Vocal messages** • Phonetics and phonology • Prosody and voice quality • Vocal expressions • Audio representation and basic feature extraction • Praat and OpenEar	**Introduction** • Morency et al. (2010), Human Communication dynamics • Vinciarelli et al. (2009), Social Signal Processing • Krauss et al. (2002), The psychology of Verbal Communication • *(optional) Pentland (2008), Honest Signals, Ch. 1*	• •
Week 3	**Visual messages** • Gesture, gaze, posture and proxemics • Facial expressions • Image and video representation • Watson, FaceAPI, AAM and EyeAPI	**Vocal messages** • Ang et al. (2002), Prosodic-based detection of annoyance and frustration • Scherer (2003), Vocal communication of emotion: A review of research paradigms • Bachorowsky et al. (2001), The acoustic features of laughter • *(optional) Ladefoged (2004), A course in phonetics* • *(optional) Jurafsky and Martin (2008), Speech and Language Processing, Ch. 7, Sect. 7.1-7.4*	•
Week 4 *** Draft project proposals*	**Study Design, Evaluation and Analysis** • User studies • Coder agreement, kappa • Statistical analysis • Student T-test, effect-size	**Visual messages** • Kendon (1995) Gesture studies • Kramer (2008) Nonverbal communication • Tian et al. (2006) Facial expression	• •

due.	• Wisdom of crowd	analysis • (optional) Argyle and Dean (1965) Eye-Contact, Distance and Affiliation	
Week 5 *** First homework due*	**Machine Learning: basic concepts** • Classification and Evaluation methods • Training/validation and testing • Decision models • Support Vector Machine • ML toolbox (weka)	**Study Design, Evaluation and Analysis** • Leroy (2011), Designing User Studies in Informatics • Krippendorff (2011), Agreement and Information in the Reliability of Coding • TBD	•
Week 6	**Behavior Analysis and unsupervised learning** • Dimensionality reduction • Data clustering • Dynamic time warping • Feature selection	**Machine Learning: basic concepts** • Fawcett (2006) An introduction to ROC analysis • Christoudias et al. (2006) Co-adaptation of audio-visual speech and gestures • (optional) Langley and Kibler (1991), The Experimental Study of Machine Learning	•
Week 7 *** Project proposals due.*	**Verbal messages** • Language models and N-grams • Boundaries, fillers and disfluencies • Syntax and part-of-speech tagging • Sphinx, hTK and syntax parsers	**Behavior Analysis and unsupervised learning** • Wu et al. (2004) Multimedia data analysis • Zhou et al. (2010) Unsupervised discovery of facial events	•
Week 8 *** Second homework due*	**Conversational messages** • Discourse analysis • Turn-taking and backchannel • Semantics and pragmatics • Speech and dialogue acts	**Verbal messages** • • • *(optional) Jurafsky and Martin (2008), Speech and Language Processing, 4.1-4.4, 5.1-5.3 and 12.1-12.2* • *(optional) Kim and Hovy (2004) Determining the sentiment of opinions* • *(optional) Liu et al. (2004) Metadata extraction*	•

Week 9	**Affective messages and personality traits** • Emotion and cognitive modeling • Big five personality dimensions • Social behaviors	**Conversational messages** • Duncan (1974) Signals for speaking turns • Stolcke et al (2000) Dialogue act modeling • Bohus and Horvitz (2010), Computational Turn-taking • *(optional) Jurafsky and Martin (2008), Speech and Language Processing, Sect. 17.2-17.3 and 21.1-21.4* • *(optional) Clark and Brennan (1991) Grounding in Communication*	•
Week 10	**Multimodal behavior recognition (1/3)** • Fuzzy machine learning • Neural networks • Deep belief networks	**Affective messages and personality traits** • • • *(optional) Gratch and Marsella (2005), Emotion Psychology* • *(optional) Barrick and Mount (1991), Big Five personality*	•
Week 11 *** Mid-term reports due.*	**Multimodal behavior recognition (2/3)** • Multimodal fusion • Audio-visual recognition • Hidden Markov Models • Multi-streams, coupled, factorial and asynchronous HMMs	**Multimodal behavior recognition (1/3)** • Multimodal Deep Learning • Scherer et al. (2012), Spotting Laughter in naturalistic multiparty conversations: a comparison of automatic online and offline approaches using audiovisual data • *(optional) McNeill (1985) Gestures* • *(optional) Scherer et al. (2012), Investigating Fuzzy-Input Fuzzy-Output Support Vector Machines for Robust Voice Quality Classification* • *(optional) Ambady and Rosenthal (1992) Thin slicing*	•

		• *(optional) P. Verlinde and G. Chollet (1999) Decision fusion paradigms*	
Week 12	**Multimodal behavior recognition (3/3)** • Maximum entropy models • Conditional random fields • Latent-dynamic CRF • Multi-view HCRF model	**Multimodal behavior recognition (2/3)** • McCowan et al (2005) Multimodal group actions • *(optional) Gros et al. (2008), Multimodal processing and Interaction, Chapter 1* • *(optional) Atrey et al (2010) Multimodal fusion for multimedia analysis: a survey* • *(optional) Nefian et al. (2002) Audio-visual speech recognition* • *(optional) Christoudias (2006) Co-adaptation*	•
Week 13	**Thanksgiving**		
Week 14 *** Demos and lecture at ICT*	**Multimodal applications and demos** • Multimodal interaction • Live ICT demos • Gunslinger demo • MultiSense demo • Lightstage demo	**Multimodal behavior recognition (3/3)** • Bousmalis et al (2011) Modeling hidden dynamics of agreement/disagreement • *(optional) El Kaliouby and Robinson (2005) Real-Time Inference of Complex Mental States* • *(optional) Morency et al. (2008) Context-based recognition* • *(optional) Morency et al (2007) Latent-dynamic CRF* • *(optional) Tong et al. (2009) A unified probabilistic framework for facial action modeling*	•
Week 15 *** Final projects due Sunday 12/9*	**Final project presentations**		

at 11:59pm

Bibliography

Primary readings

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 [USC blackboard]

Vocal messages

5. Ang, Jeremy ; Dhillon, Rajdip ; Krupski, Ashley ; Shriberg, Elizabeth ; Stolcke, Andreas (2002): Prosody-based automatic detection of annoyance and frustration in human-computer dialog, In *ICSLP-2002*, 2037-2040
6. Scherer (2003), Vocal communication of emotion: A review of research paradigms, *Speech Communication*, 40, pp. 227-256
7. Bachorowsky et al. (2001), The acoustic features of laughter, *Journal of the Acoustic Society of America*, 110 (3), pp. 1581-1597
8. (Optional) Ladefoged (2004), A course in phonetics [USC blackboard]
9. (Optional) Jurafsky and Martin (2008), Speech and Language Processing, Chapter 7, Sections 7.1-7.4 [USC blackboard]

Visual messages

10. Adam Kendon, An Agenda for Gesture Studies, This article appeared in Volume 7 (3) of the Semiotic Review of Books.
11. 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. [USC blackboard]
12. Ying-Li Tian, Takeo Kanade and Jeffrey F. Cohn, Facial Expression Analysis, *Handbook of Face Recognition*, 2005, 247-275
13. (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

14. Leroy (2011), Designing User Studies in Informatics, Springer.
15. Krippendorff (2011), Agreement and Information in the Reliability of Coding, *Communication Methods and Measures*, 5 (2), pp. 93-112
16. Kingstone et al. (2011), Eyes Wide Shut, *Attention, Perception and Psychophysics*, Springer, pp. 291-296.

Machine Learning: Basic Concepts

17. Tom Fawcett, An introduction to ROC analysis, *Pattern Recognition Letters,* Volume 27, Issue 8, June 2006, Pages 861–874
18. 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
19. (optional) Langley and Kibler (1991), The Experimental Study of Machine Learning, Unpublished paper.

Behavior analysis and unsupervised learning

20. Y.Wu, E. Chang, K. Chang and J. Smith (2004) Optimal multimodal fusion for multimedia data analysis, Proceedings of the 12th annual ACM international conference on Multimedia (ACM Multimedia)
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)

Verbal messages

22.
23.
24. (optional) Jurafsky and Martin (2008), Speech and Language Processing, Sections 4.1-4.4, 5.1-5.3 and 12.1-12.2 [USC blackboard]
25. (optional) Soo-Min Kim and Eduard Hovy (2004) Determining the Sentiment of Opinions, Proceedings of the COLING conference, Geneva
26. (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

27. Duncan (1974) Some Signals and Rules for Taking Speaking Turns in Conversations
28. 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
29. Bohus, D., Horvitz, E., (2010) - Computational Models for Multiparty Turn-Taking, Microsoft Technical Report MSR-TR-2010-115
30. (optional) Jurafsky and Martin (2008), Speech and Language Processing, Sections 17.1-17.4 and 21.1-21.4 [USC blackboard]
31. (optional) Clark and Brennan (1991) Grounding in Communication, [USC blackboard]

Affective messages and personality traits

32.
33.
34. (optional) Gratch & Marsella, 2005 Lessons from Emotion Psychology for the Design of Lifelike Characters
35. (optional) Mr Barrick, Mk Mount (1991) The Big Five Personality Dimensions And Job Performance: A Meta-Analysis - Personnel Psychology

Multimodal behavior recognition (1/3)

36.

37. 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.
38. (optional) McNeill, D. (1985). "So you think gestures are nonverbal?" In *Psychological Review* 92:350-371.
39. (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.
40. (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
41. (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

Multimodal behavior recognition (2/3)

42. 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
43. (optional) Gros, Potamianos and Maragos (2008) Multimodal Processing and Interaction, SpringerLink, Chapter 1 [SpringerLink or USC blackboard]
44. (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
45. (optional) 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

Multimodal behavior recognition (3/3)

46. Konstantinos Bousmalis, Louis–Philippe Morency and Maja Pantic, Modeling Hidden Dynamics of Multimodal Cues for Spontaneous Agreement and Disagreement Recognition, Face and Gestures 2011
47. (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
48. (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.*
49. (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

Grades

- **Grading breakdown**
  - Attendance and participation 10% (1 free absence)
  - Reading assignments 15%
  - Leading class discussion 15%
  - Two practical exercises 20% (10% each)
  - Course project:
    - Proposal and mid-term report 15%
    - final report and presentation 25%

- **Attendance**
  - Students are expected to attend every class (1 free absence allowed) and participate actively during the group discussions.
- **Reading assignments**
  - The reading assignment for each class will consist of 2-4 research papers (posted online at least one week before the class). These papers are specially selected to complement the lectures and show state-of-the-are research.
  - Sunday before each class, 1-3 questions will be posted online.
  - Students must send their answers by 5pm the day before the class. The answers will be part of the group discussion.
- **Group discussions**
  - Each student will be leading the group discussion twice during the semester. A signup sheet will be available during the first class.
  - Students can lead the discussion individually or pair with another student. The pairing should be different for the second group discussion.
  - 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.
- **Practical exercises**
  - These two practical exercises will be designed to give hands-on experience with machine learning (e.g., SVM, HMM, CRF) for multimodal behavior recognition.
  - Each practical will come with sample code (Matlab) and links to existing machine learning libraries.
  - Students will need to submit their code (zip files) with their answer to each practical exercise.
- **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 team of two. The mid-term and final report will need to outline the tasks of each participant. Team projects will be expected to include a deeper analysis than individual projects.
- **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 language, vocal and nonverbal 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.

Statement for Students with Disabilities

Any student requesting academic accommodations based on a disability is required to register with Disability Services and Programs (DSP) each semester. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered to me (or to TA) as early in the semester as possible. DSP is located in STU 301 and is open 8:30 a.m.–5:00 p.m., Monday through Friday. The phone number for DSP is (213) 740-0776.

Statement on Academic Integrity

USC seeks to maintain an optimal learning environment. General principles of academic honesty include the concept of respect for the intellectual property of others, the expectation that individual work will be submitted unless otherwise allowed by an instructor, and the obligations both to protect one's own academic work from misuse by others as well as to avoid using another's work as one's own. All students are expected to understand and abide by these principles. *Scampus,* the Student Guidebook, contains the Student Conduct Code in Section 11.00, while the recommended sanctions are located in Appendix A: http://www.usc.edu/dept/publications/SCAMPUS/gov/. Students will be referred to the Office of Student Judicial Affairs and Community Standards for further review, should there be any suspicion of academic dishonesty. The Review process can be found at: http://www.usc.edu/student-affairs/SJACS/.