Human Communication and Machine Learning (Fall 2010)
Course description and lessons learned
Louis-Philippe Morency
Talk Outline

- Course on Human Communication and Machine Learning
  - Related courses
  - Scientific areas
  - Learning goals and prerequisites
  - Books and literature
  - Course outline
  - Course projects
  - Useful links
Related Courses

- Justine Cassell (Northwestern University):
  - Theories and Technologies for Human Communication

- Dan Jurafski (Stanford):
  - Speech Recognition, Synthesis, and Dialogue

- Rosalind Picard (MIT):
  - Affective Computing

- Carlos Busso (UT Dallas):
  - Multimodal Signal Processing

- Stacy Marsella and Jonathan Gratch (USC):
  - Affective Computing
Scientific Areas

- Human Communication
  - Verbal and nonverbal Communication
  - Interpersonal and small group Communication
- Computer vision and signal processing
- Machine learning and artificial intelligence
- Computational linguistics and dialog modeling
- Emotion and cognitive sciences
- Human-computer interaction
  - including human-robot interaction and virtual humans
Course Facts

- 2-3 months of preparation (1 week in London)
- 22 graduate students registered
- About 50% PhD and 50% Master students
- 3 hours per week for 15 weeks
- 2 hours of lecture + 1 hour of paper discussion
- 15 course projects
Course Objectives

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;
Course Objectives

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.
Recommended Preparation

- Suggested preparation courses:
  - CSCI 542: Neural Computation with Artificial Neural Networks
  - CSCI 567: Machine Learning
  - CSCI 573: Advanced Artificial Intelligence

- Academic background
  - Probability and statistic
  - Linear algebra
  - Matlab

- This course is not a replacement for the Machine Learning course (CSCI 567).
Related Textbooks

- *Nonverbal Communication in Human Interaction (7th edition)*, Knapp and Hall, Wadsworth, 2010
- *Speech and Language Processing (2nd edition)*, Jurafsky and Martin, Pearson, 2008
- *Text-to-speech Synthesis*, Taylor, 2009
# Course Outline: Introduction

## Lectures (2:00pm-3:50pm)

<table>
<thead>
<tr>
<th>Introduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>• A multi-modal, multi-party, multi-label dynamic problem</td>
</tr>
<tr>
<td>• Human communication dynamics</td>
</tr>
<tr>
<td>• Applications and domains</td>
</tr>
<tr>
<td>• Mid-term and final projects</td>
</tr>
</tbody>
</table>

## Readings for discussion sessions (4:00pm-4:50pm)

<table>
<thead>
<tr>
<th>Introduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Morency et al. (2010), Human Communication dynamics</td>
</tr>
<tr>
<td>• Vinciarelli et al. (2009), Social Signal Processing</td>
</tr>
<tr>
<td>• Carletta (2007), AMI dataset</td>
</tr>
</tbody>
</table>

## Communication models

<table>
<thead>
<tr>
<th>Communication models</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Emitter-receiver models</td>
</tr>
<tr>
<td>• Communicative signals: signs and symbols</td>
</tr>
<tr>
<td>• Common ground</td>
</tr>
<tr>
<td>• Datasets and sensing tools</td>
</tr>
</tbody>
</table>

## Communication models

<table>
<thead>
<tr>
<th>Communication models</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Krauss et al. (2002), The psychology of Verbal Communication</td>
</tr>
<tr>
<td>• Clark and Brennan (1991) Grounding in Communication</td>
</tr>
<tr>
<td>• Pentland (2008), Honest Signals, Ch. 1</td>
</tr>
<tr>
<td>• (optional) Taylor (2009) Text-to-speech Synthesis, Chapter 2</td>
</tr>
</tbody>
</table>
# Course Outline: Communicative Messages

<table>
<thead>
<tr>
<th>Verbal messages</th>
<th>Verbal messages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Language models and N-grams</td>
<td>• Jurafsky and Martin (2008), Speech and Language Processing, 4.1-4.4, 5.1-5.3 and 12.1-12.2</td>
</tr>
<tr>
<td>• Boundaries, fillers and disfluencies</td>
<td>• Kim and Hovy (2004) Determining the sentiment of opinions</td>
</tr>
<tr>
<td>• Syntax and part-of-speech tagging</td>
<td>• Liu et al. (2004) Metadata extraction</td>
</tr>
<tr>
<td>• Sphinx, hTK and syntax parsers</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vocal messages</th>
<th>Vocal messages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Phonetics and phonology</td>
<td>• Taylor (2009), Text-to-speech Synthesis, Sections 6.1-6.5 and 9.1-9.2</td>
</tr>
<tr>
<td>• Rhythm, stress and Intonation</td>
<td>• Ang et al. (2002), Prosodic-based detection of annoyance and frustration</td>
</tr>
<tr>
<td>• Audio representation</td>
<td>• Ward and Tsukahara (2000)</td>
</tr>
<tr>
<td>• Praat and OpenEar</td>
<td>• (optional) Jurafsky and Martin (2008), Speech and Language Processing, Ch. 7, Sect. 7.1-7.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Visual messages</th>
<th>Visual messages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Gesture, gaze, posture and proxemics</td>
<td>• Kramer (2008) Nonverbal communication</td>
</tr>
<tr>
<td>• Facial expressions</td>
<td>• Kendon (1995) Gesture studies</td>
</tr>
<tr>
<td>• Image and video representation</td>
<td>• Argyle and Dean (1965) Eye-Contact, Distance and Affiliation</td>
</tr>
<tr>
<td>• Watson, FaceAPI, AAM and EyeAPI</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conversational messages</th>
<th>Conversational messages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Discourse analysis</td>
<td>• Duncan (1974) Signals for speaking turns</td>
</tr>
<tr>
<td>• Turn-taking and backchannel</td>
<td>• Stolcke et al (2000) Dialogue act modeling</td>
</tr>
<tr>
<td>• Semantics and pragmatics</td>
<td>• Bohus and Horvitz (2010), Computational Turn-taking</td>
</tr>
<tr>
<td>• Speech and dialogue acts</td>
<td>• (optional) Jurafsky and Martin (2008), Speech and Language Processing</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Affective messages and personality traits</th>
<th>Affective messages and personality traits</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Emotion and cognitive modeling</td>
<td>• Gratch and Marsella (2005), Emotion Psychology</td>
</tr>
<tr>
<td>• Big five personality dimensions</td>
<td>• Barrick and Mount (1991), Big Five personality</td>
</tr>
<tr>
<td>• Social behaviors</td>
<td></td>
</tr>
</tbody>
</table>
# Course Outline: Multimodal Recognition

## Multimodal representation
- Speech and gestures
- Signals and symbols
- Multimodal fusion
- Statistical analysis

## Multimodal processing
- Audio-visual recognition
- Hidden Markov Models
- Multi-streams, coupled, factorial and asynchronous HMMs

## Multimodal Behavior analysis
- Dimensionality reduction
- Data clustering
- Dynamic time warping
- Feature selection

## Multimodal behavior recognition (1/2)
- Bootstrapping and Co-training
- Nearest-neighbor
- Decision trees
- Support vector machines

## Multimodal behavior recognition (1/2)
- Christoudias et al. (2006) Co-adaptation of audio-visual speech and gestures
- Kapoor and Picard (2005) Multimodal affect recognition
- (optional) P. Verlinde and G. Chollet (1999) Decision fusion paradigms
- (optional) Chapter 9 of Machine Learning for Audio, Image and Video

## Behavior recognition (2/2)
- Conditional random fields
- Latent-dynamic CRF
- Dynamic Bayesian networks

## Behavior recognition (2/2)
- El Kaliouby and Robinson (2005) Real-Time Inference of Complex Mental States
- (optional) Tong et al. (2009) A unified probabilistic framework for facial action modeling
# Course Outline: Evaluation

<table>
<thead>
<tr>
<th>Lectures</th>
<th>Readings for discussion sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2:00pm-3:50pm)</td>
<td>(4:00pm-4:50pm)</td>
</tr>
<tr>
<td><strong>Subjective and quantitative evaluations</strong></td>
<td></td>
</tr>
<tr>
<td>- Coder agreement, kappa</td>
<td></td>
</tr>
<tr>
<td>- User studies</td>
<td></td>
</tr>
</tbody>
</table>

Final project presentations
Group Discussions

- Each student will be leading the group discussion twice during the semester.
  - Students can lead the discussion individually or pair with another student.
  - The pairing should be different for the second group discussion.

- The discussion should bring something new and interactive to the class.
  - Simple implementation of the algorithms,
  - New challenging questions and applications,
  - Example datasets.
Course Project

- **Project goals**
  - analyze human communicative behaviors in social settings using state-of-the-art statistical and probabilistic models.
  - The course project is specifically designed to give students practical experience in computational study of human social communication.
Course Project

• **Mid-term report**
  
  – Present a qualitative analysis of the selected dataset and communicative behaviors.
  
  – Include correct transcription and annotations of the language, vocal and nonverbal behaviors.
  
  – Using standard statistical tools and qualitative observations: highlight the challenges with this dataset (and communicative behaviors) and suggest an approach to solve them.
Course Project

- 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.
Examples of Course Projects

- Modeling an Interaction Between Children with Autism and a Humanoid Robot
- Toward a Predictive Model for Human Spatial Dynamics in Social Interaction
- A Multimodal Approach to Detecting the Mental State of Person in a Dyadic Human Conversation
- In the Footsteps of Iago: How Behaviors In One-on-One Conversations Influence Suspicion Responses
- Analysis and Recognition of Sarcasm using Audio-Visual Cues
- Genre Detection Of Video Clips
- Kinecting Body Language Cues to Signals of Engagement
Dataset Resources

- Humaine Database Wiki
  - http://emotion-research.net/wiki/Databases

- Semaine corpus
  - http://semaine-db.eu/

- Multimodal Corpora
  - http://www.multimodal-corpora.org/

- CMU Kitchen Dataset
  - http://kitchen.cs.cmu.edu/

- Prometheus database

- ICT Rapport dataset
  - http://people.ict.usc.edu/~gratch/ (new link to come)
Dataset Resources

- **ELRA Multimodal/Multimedia Resources**

- **Multimodal Corpora Information page (U Twente)**
  - http://wwwwhome.cs.utwente.nl/~zsofi/eeca/MultimodalCorporaResources.htm

- **Linguistic Data Consortium**
  - http://www.ldc.upenn.edu/

- **Language Resources and Evaluation Conferences**
  - http://www.lrec-conf.org/

- **European Language Resource Distribution Agency**
  - http://www.elda.org/

- **Switchboard**
Annotation tools

- **Transcriber**

- **ANVIL**
  - [http://www.anvil-software.de/](http://www.anvil-software.de/)

- **ELAN**
Automatic Annotations (Speech)

- **Speech recognizer:**
  - Sphinx
  - HTK
    - [http://htk.eng.cam.ac.uk/](http://htk.eng.cam.ac.uk/)

- **Language Modeling Toolkit:**
  - SRILM

- **Syntactic parser:**
  - LRDEP
    - [http://people.ict.usc.edu/~sagae/parser/](http://people.ict.usc.edu/~sagae/parser/)

- **Prosody**
  - Praat
    - [http://www.fon.hum.uva.nl/praat/](http://www.fon.hum.uva.nl/praat/)
  - OpenEar
    - [http://sourceforge.net/projects/openart/](http://sourceforge.net/projects/openart/)
Automatic Annotations (Vision)

- Facial analysis
  - Watson
    - [http://sourceforge.net/projects/watson/](http://sourceforge.net/projects/watson/)
  - FaceAPI
  - AAM
    - [http://www2.imm.dtu.dk/~aam/](http://www2.imm.dtu.dk/~aam/)
  - EyeAPI
    - [http://staff.science.uva.nl/~rvalenti/](http://staff.science.uva.nl/~rvalenti/)
Machine Learning Toolboxes

- Weka: Data mining
- LibSVM: Support vector machine
- HCRF library: Conditional random fields
- Bayes net toolbox: Bayesian networks
- Mallet
- Torch