USC Institute for Creative Technologies
University of Southern California

Modeling Human Communication Dynamics

Louis-Philippe Morency

USC Multimodal Communication and Machine Learning Lab [MultiComp Lab]

PhD students: **Derya Ozkan and Sunghyun Park**
Master student: **Moitreya Chatterjee**
Post-doctoral researcher: **Stefan Scherer**
Research programmer: **Giota Stratou**
Project manager: **Alesia Egan**

Multimodal Communication and Machine Learning Lab

© Keith Schaffer

Multimodal Perception of User State: Applications

Medical

Distress Indicators

with MIT and CogitoHealth

Suicide prevention

with Cincinnati Hospital

Education

Group learning analytics

with Stanford and UCSD

Virtual Learning Peer

with CMU

Business

Negotiation outcomes

with USC business school

YouTube: Opinion mining

with UNT and UT Dallas

Disorders

- Depression
- PTSD
- Distress

Social

- Engagement
- Dominance
- Empathy

Affect

- Frustration
- Agreement
- Sentiment

Multimodal Perception of Distress Indicators

Multimodal Communicative Behaviors

Verbal

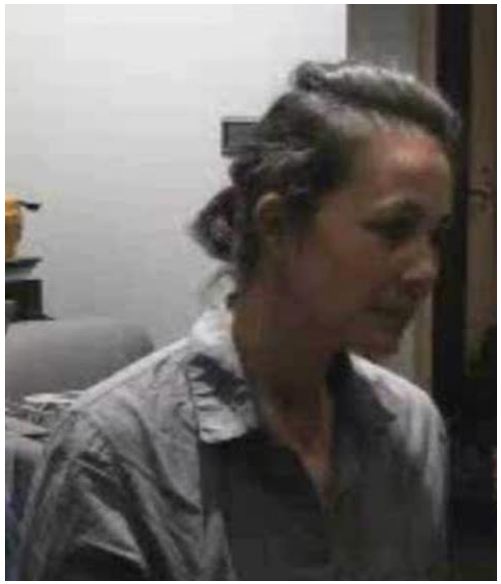
- **Lexicon**
  - Words
- **Syntax**
  - Part-of-speech
  - Dependencies
- **Pragmatics**
  - Discourse acts

Auditory

- **Prosody**
  - Intonation
  - Voice quality
- **Vocal expressions**
  - Laughter, moans

Visual

- **Gestures**
  - Head gestures
  - Eye gestures
  - Arm gestures
- **Body language**
  - Body posture
  - Proxemics
- **Eye contact**
  - Head gaze
  - Eye gaze
- **Facial expressions**
  - FACS action units
  - Smile, frowning



From Audio-Visual Signals to Perceived User State

Verbal

- **Lexicon**
  - Words
- **Syntax**
  - Part-of-speech
  - Dependencies
- **Pragmatics**
  - Discourse acts

Auditory

- **Prosody**
  - Intonation
  - Voice quality
- **Vocal expressions**
  - Laughter, moans

Visual

- **Gestures**
  - Head gestures
  - Eye gestures
  - Arm gestures
- **Body language**
  - Body posture
  - Proxemics
- **Eye contact**
  - Head gaze
  - Eye gaze
- **Facial expressions**
  - FACS action units
  - Smile, frowning

Disorders

- Depression
- PTSD
- Distress

Social

- Engagement
- Dominance
- Empathy

Affect

- Frustration
- Agreement
- Sentiment

From Audio-Visual Signals to Perceived User State

Audio signals

- Voice pitch

Visual signals

- Head pose

Perceived User State

- Distress
- Engagement
- Sentiment

Sensing → **Understanding**

From Audio-Visual Signals to Perceived User State

Audio signals
- Voice pitch

Visual signals
- Head pose

Sensing

KINECT for XBOX 360

- Low-cost depth sensor
- Articulated body tracking

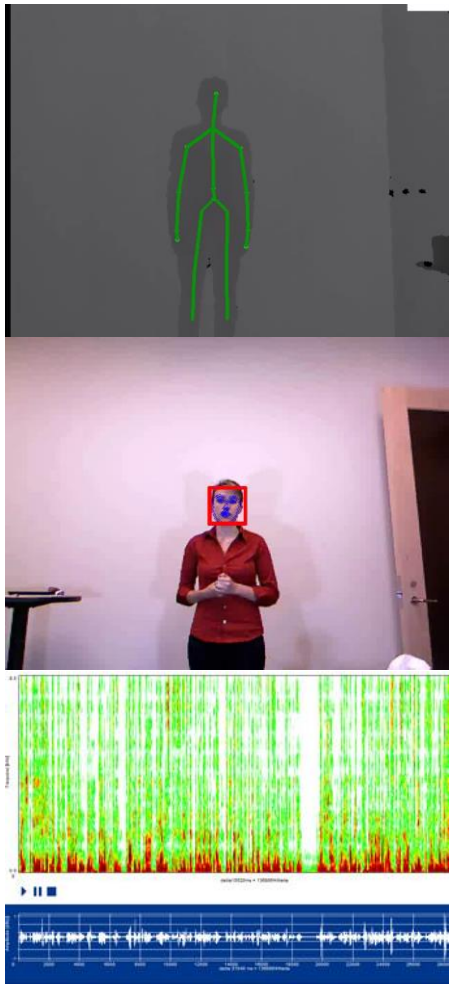
ICT **GAVAM**
[FG 2008, best paper award]

CLM-Z [CVPR 2012]

- 3D head pose estimation
- Real-time facial feature tracker

ICT **Audio Toolbox**

- Pitch, energy and speaking rate
- Automatic voice quality analysis



From Audio-Visual Signals to Perceived User State

Audio signals
- Voice pitch

Visual signals
- Head pose

Perceived User State
- Distress
- Engagement
- Sentiment

Sensing → **Understanding**

Behaviors

Multimodal

Interaction

Cultural

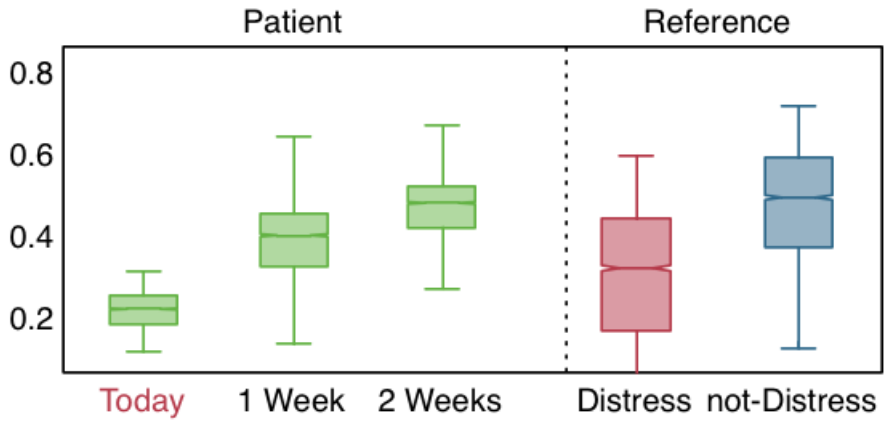
Human Communication Dynamics

Computational Behavior Indicators

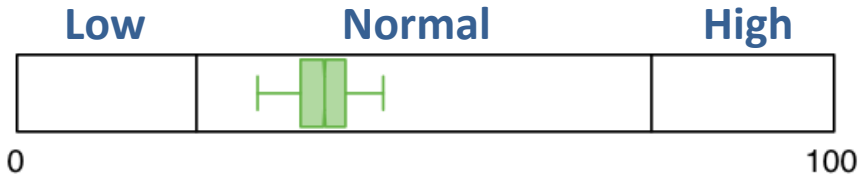
- Human in the loop:
  - Identify behaviors useful to human task
- Behavior quantification
  - Quantify changes in human behaviors

Visualization of Computational Behavior Indicators

- Transparent comparisons of observed behavioral indicators



- Analogous to medical lab result sheets



Detection and Computational Analysis of Psychological Signals (DCAPS)

Albert (Skip) Rizzo
Clinical Expert

Louis-Philippe Morency
Multimodal Perception

Jonathan Gratch
Evaluation

Arno Hartholt
Integration

Stacy Marsella
Animation

David Traum
Dialogue

cogito

Visualization & Audio Analysis

+ 27 researchers, programmers, artists and clinicians

Psychological Distress: Datasets

Aims:

- Study behaviors associated to psychological distress
  - Depression (PHQ-9)
  - Trait Anxiety (STAI)
  - PTSD (PCL-C/PCL-M)
- Examine how these cues may differ
  - Across different interaction settings
    - Face-to-face: expected to evoke strongest indicators
    - Computer-mediated (TeleCoach): intended use case
    - Human-computer (SimSensei): intended use case
  - Across different populations
    - General Los Angeles population (Craigslist)
    - Recent veterans (US Vets)

craigslist

post to classifieds

U.S. VETS

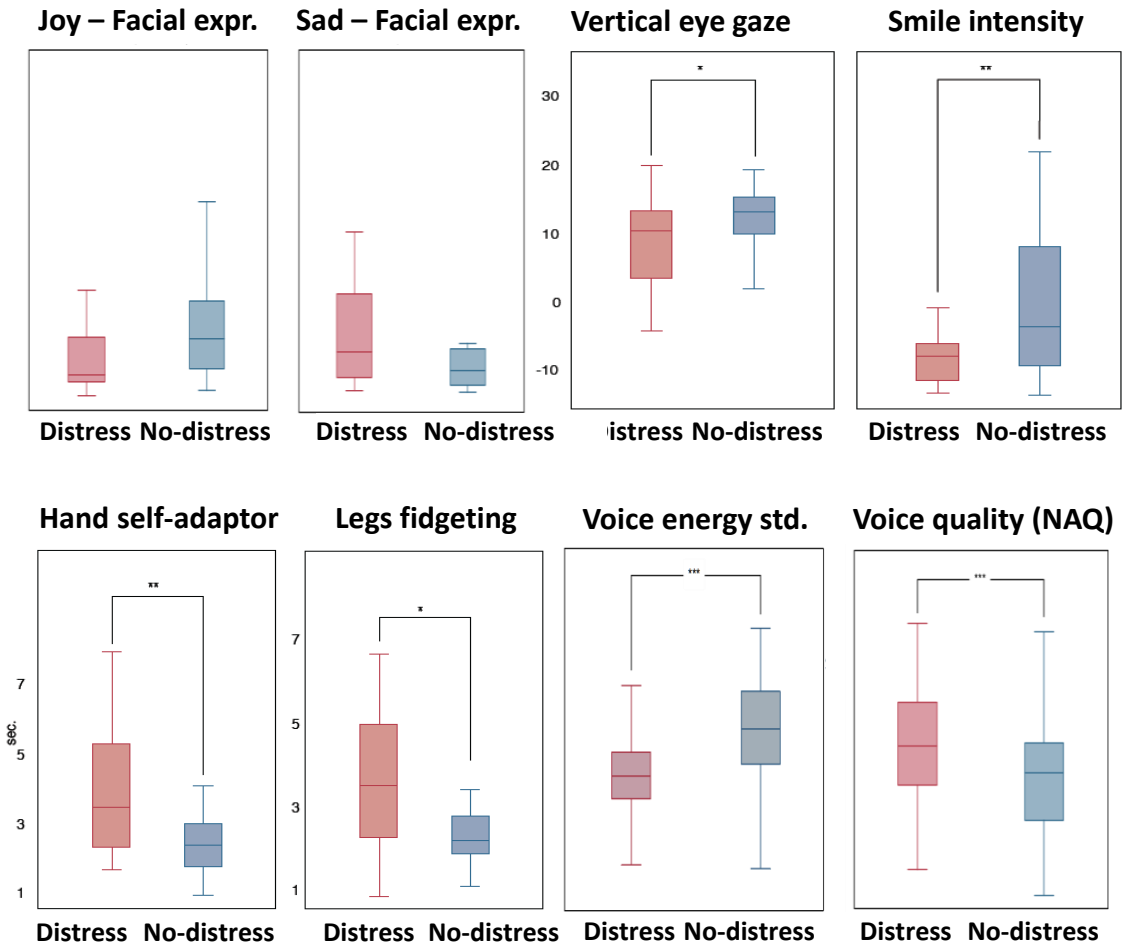
UNITED STATES VETERANS INITIATIVE

Multimodal Perception of Distress Indicators

Psychological Distress Indicators

[IEEE FG 2013 – Best paper award]

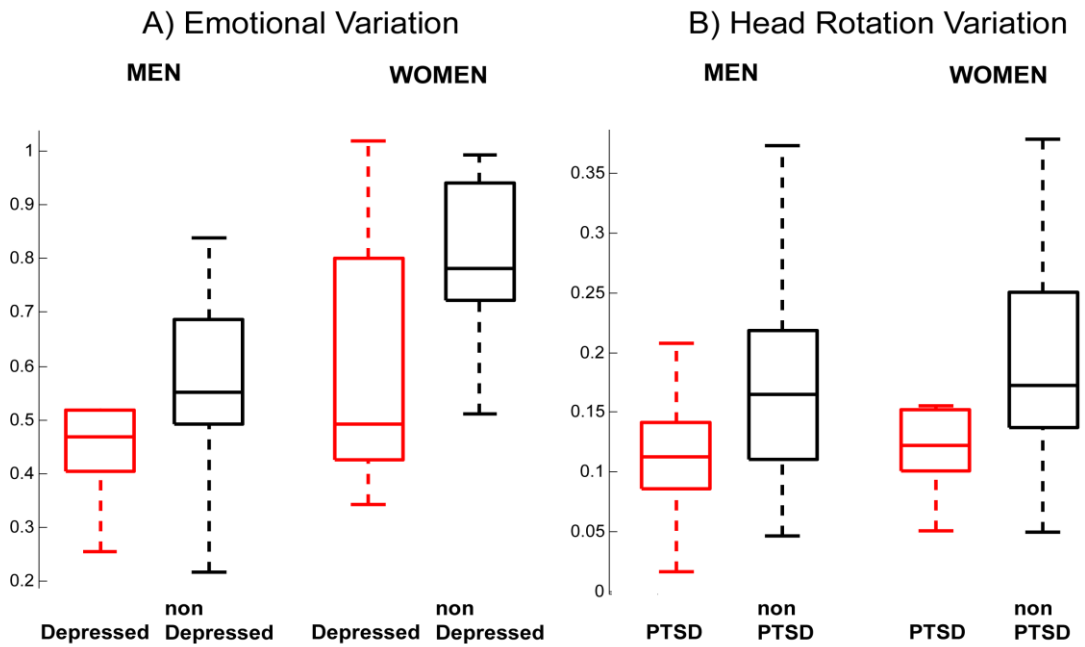
- **Distress**
- Anxiety
- Depression
- PTSD



Effect of Gender on Distress Indicators

[ACII 2013]

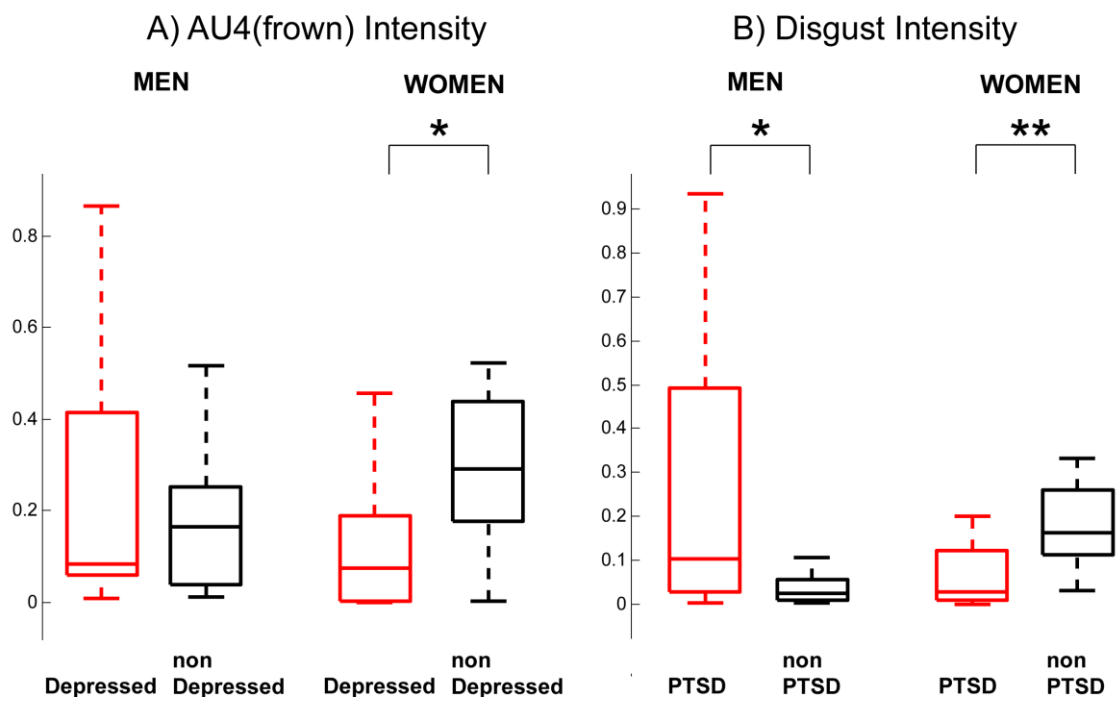
- **Distress**
- Anxiety
- Depression
- PTSD



Effect of Gender on Distress Indicators

[ACII 2013]

- **Distress**
- Anxiety
- Depression
- PTSD



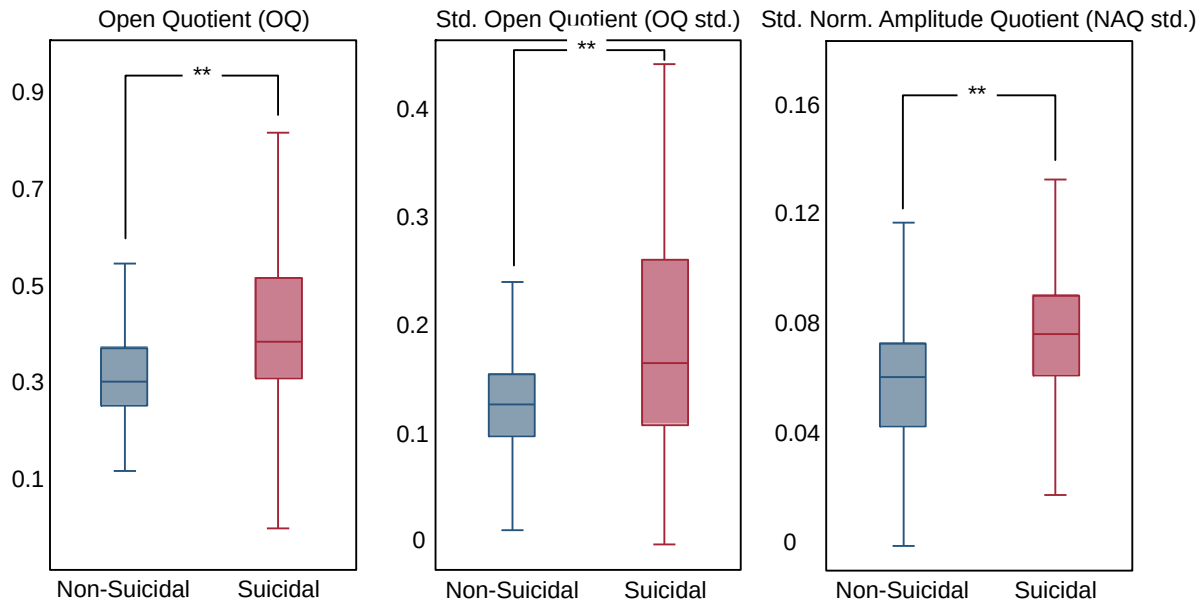
Suicide Prevention

[ICASSP 2013]

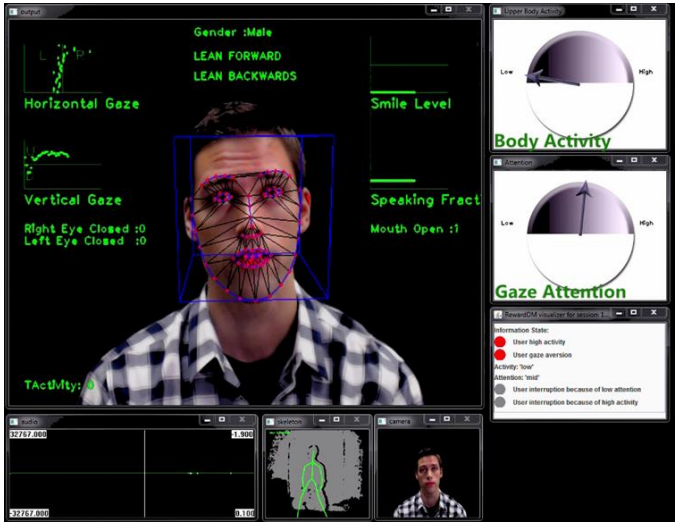
- Nonverbal indicators of suicidal ideations
- Dataset: 30 suicidal adolescents/30 non-suicidal adolescents
- Suicidal teenagers use **more breathy tones**

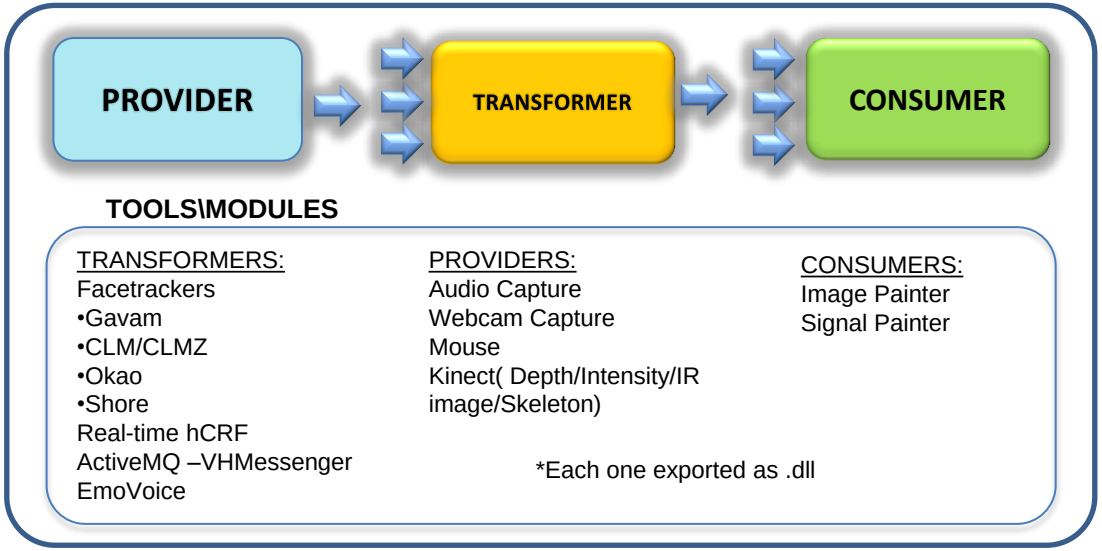
Source: CDC

	Age Groups
Rank	11-17
1	Unintentional Injury 3,924
2	Homicide 1,101
3	Suicide 823
4	Malignant Neoplasms 745
5	Heart Disease



MultiSense: Multimodal Perception Library





PML:

Perception Markup language

Sensing Layer

```
<person id="subjectA">
        <sensingLayer>
                <headPose>
                        <position z="223" y="345" x="193" />
                        <rotation rotZ="15" rotY="35" rotX="10" />
                        <confidence>0.34<confidence/>
                </headPose>
        ...
        </sensingLayer>
</person>
```

Behavior Layer

```
<person id="subjectB">
        <behaviorLayer>
                <behavior>
                        <type>attention</type>
                        <level>high</level>
                        <value>0.6</value>
                        <confidence>0.46<confidence/>
                </behavior>
                ...
        </behaviorLayer>
</person>
```

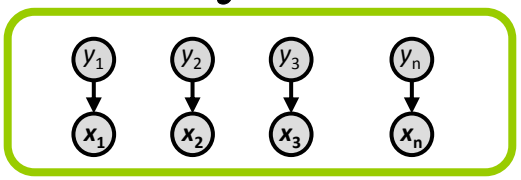
Temporal Probabilistic Learning

Multimodal

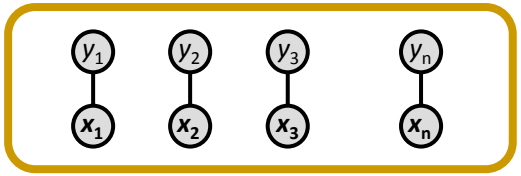
- Audio
- Visual
- Verbal

I. Temporal dynamic

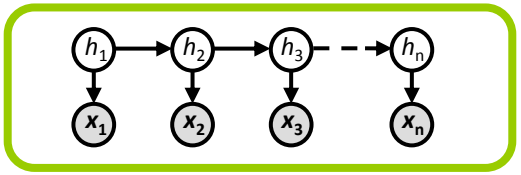
Naïve Bayes Classifier



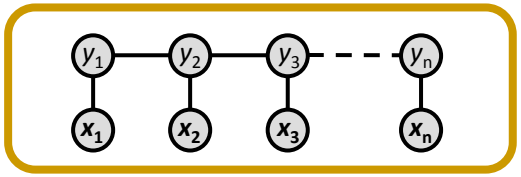
Maximum Entropy Model & Support Vector Machine



Hidden Markov Model



Conditional Random Field



Caption

y_i	Labels (e.g., `head-nod`, `other-gesture`)
x_i	Observations (e.g., `yaw, roll pitch`)
	Generative models
	Discriminative models

Temporal Probabilistic Learning

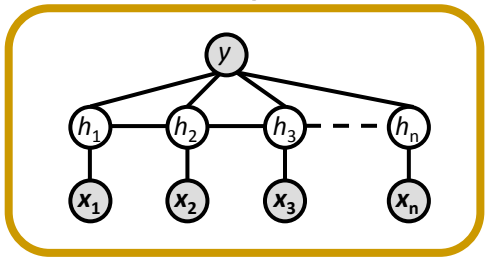
Multimodal

- Audio
- Visual
- Verbal

I. Temporal dynamic

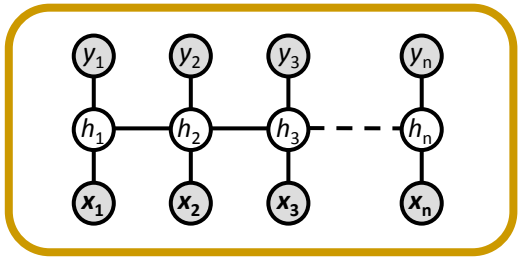
II. Hidden substructure

Hidden Conditional Random Field



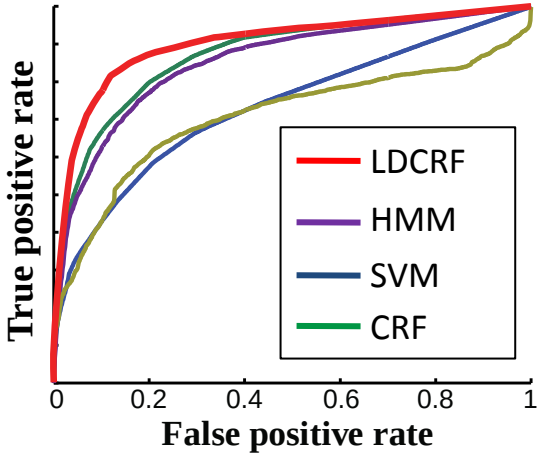
[CVPR 2006,PAMI 2007]

Latent-Dynamic Conditional Random Field



[CVPR 2007]

Model	Accuracy
HMM	84.2%
CRF	86.0%
HCRF (w=0)	91.6%
HCRF (w=1)	93.9%



Latent-Dynamic Conditional Neural Field

[FG 2013, submitted]

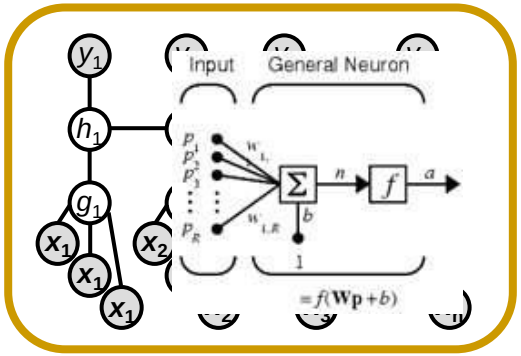
Multimodal

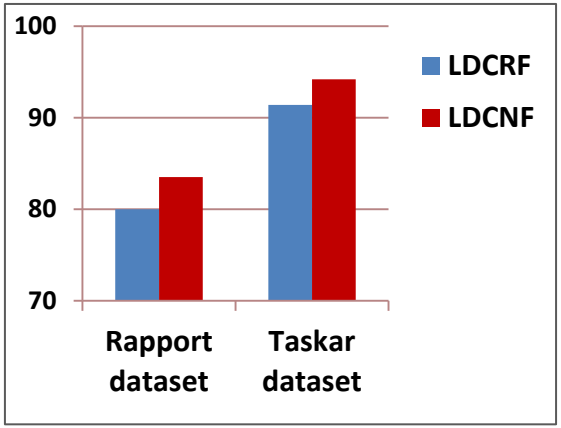
- Audio
- Visual
- Verbal

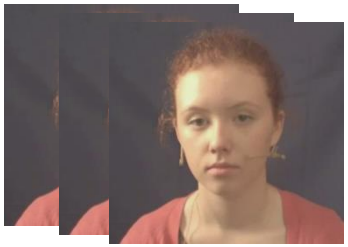
I. Temporal dynamic

II. Hidden substructure

III. Nonlinear input fusion



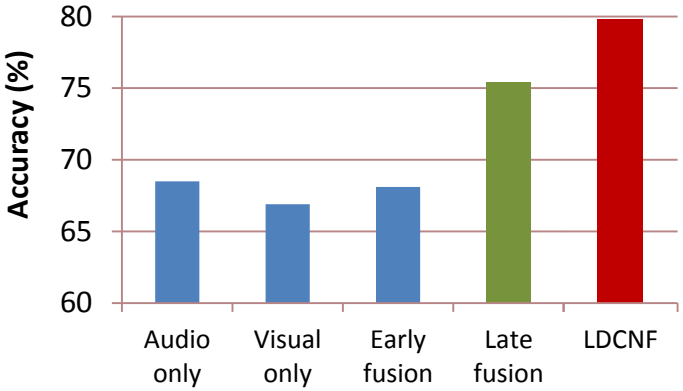




AVEC dataset

(95 videos)

Audio-visual sub-challenge



Multi-View Hidden Conditional Random Field

[CVPR 2012]

Multimodal

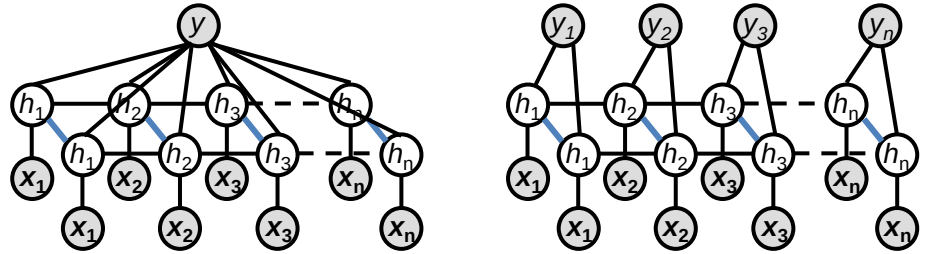
- Audio
- Visual
- Verbal

I. Temporal dynamic

II. Hidden substructure

III. Nonlinear input fusion

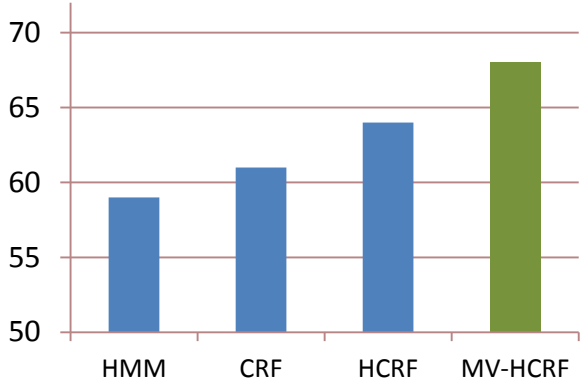
IV. Multi-stream models



Multi-View HCRF

Multi-View LDCRF

Canal 9 debate dataset



Multi-View Hidden Conditional Random Field

[ICMI 2012]

Multimodal

- Audio
- Visual
- Verbal

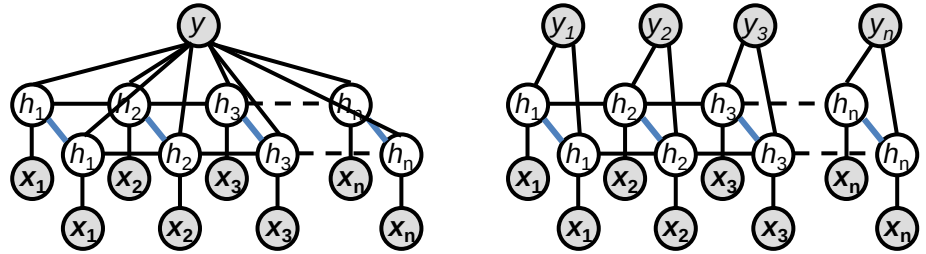
I. Temporal dynamic

II. Hidden substructure

III. Nonlinear input fusion

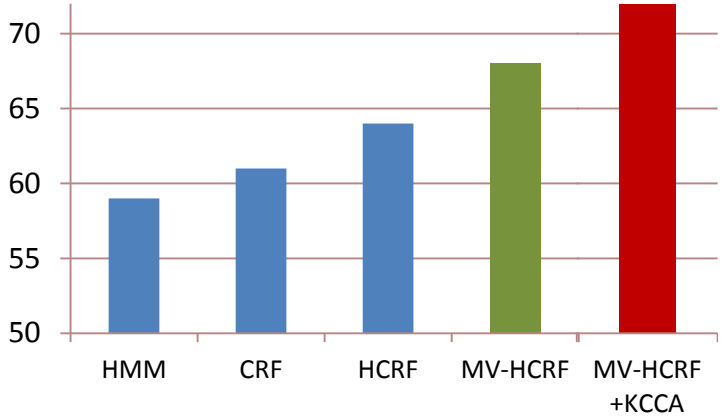
IV. Multi-stream models

V. Multimodal synchrony



Multi-View HCRF Multi-View LDCRF

Canal 9 debate dataset



Multimodal Machine Learning

Multimodal

- Audio
- Visual
- Verbal

I. Temporal dynamic in label sequence
II. Hidden substructure in label sequence
III. Nonlinear modeling of instantaneous input features
IV. Multi-stream modeling of hidden substructure
V. Synchrony in multimodal input streams
VI. Multi-label structure and correlation
VII. Symbol and signal integration
VIII. Uncertainty in behavior labels

HCRF Library:

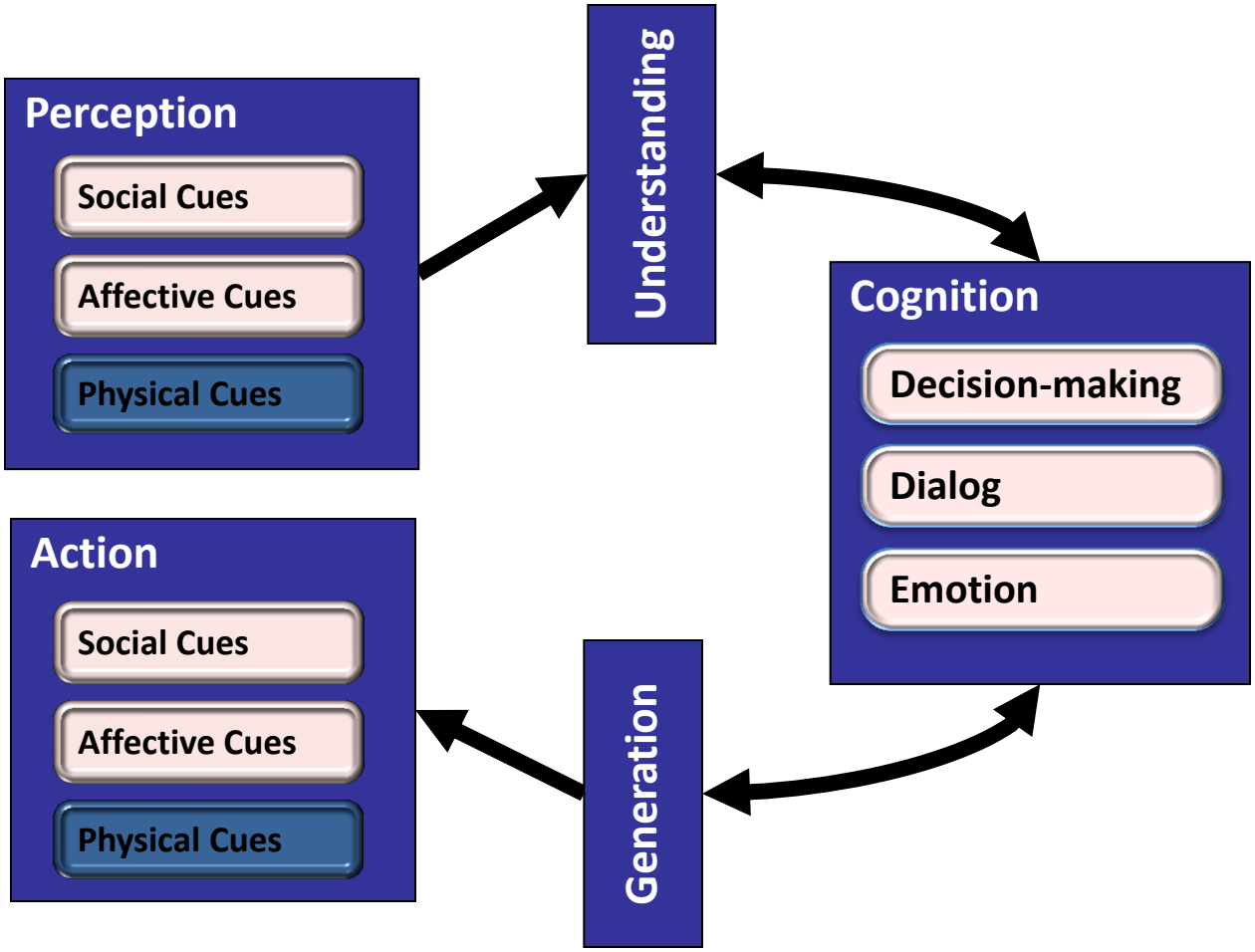
ICT open-source machine learning library

http://hcrf.sf.net

SimSensei Virtual Human Interaction Loop

Physical World

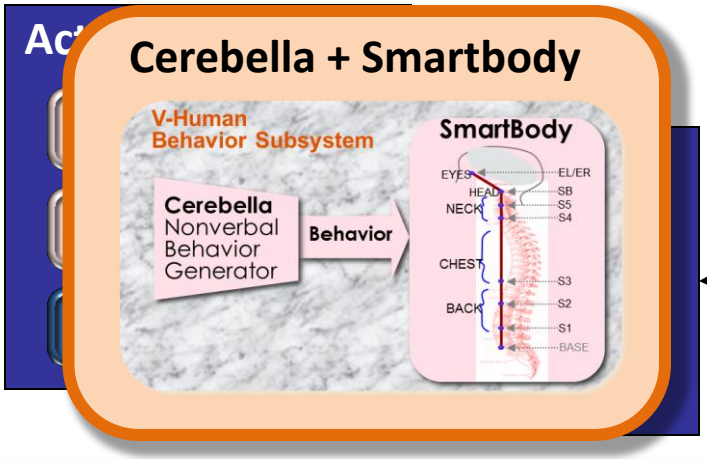
Virtual World

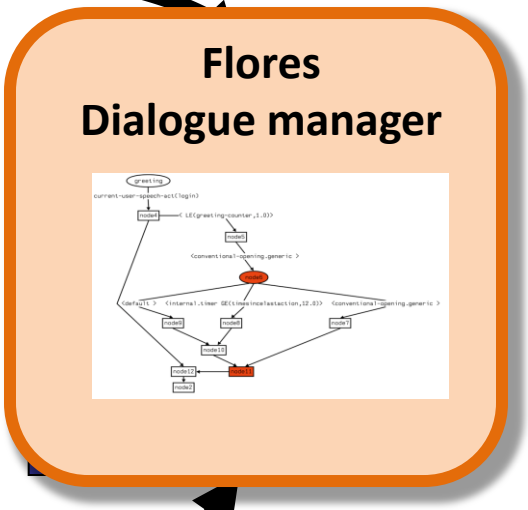


SimSensei Virtual Human Interaction Loop

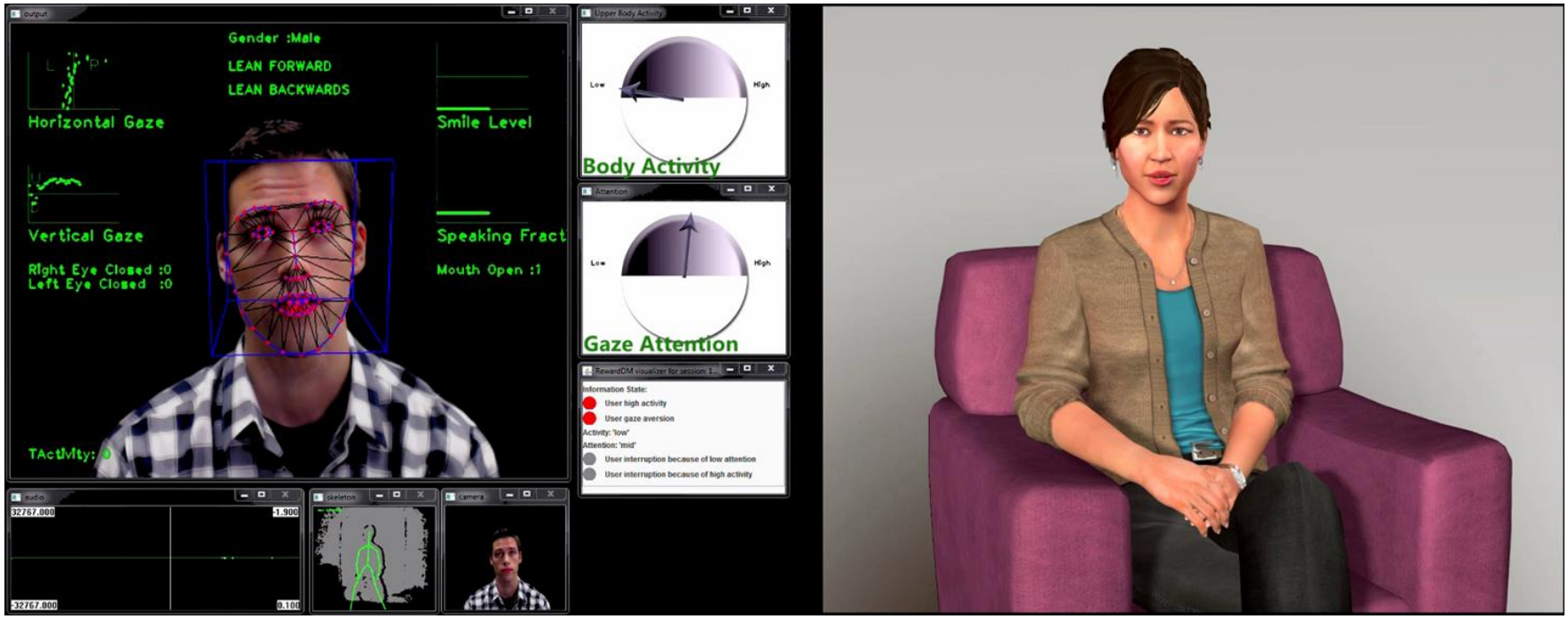
Physical World

Virtual World





MultiSense + SimSensei: Video Demonstration



USC Institute for Creative Technologies

University of Southern California

DCAPS:

Detection & Computational Analysis of Psychological Signals

3rd Interim Progress Video

Albert (Skip) Rizzo, PI
Louis-Philippe Morency, Co-PI

The work depicted here was sponsored by the U.S Defense Advanced Research Projects Agency. Statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

Collaborations and Available Technologies

- MultiSense: standardized perception framework
  - Real-time facial tracking, articulated body tracking and auditory analysis
  - Modular and multi-threaded architecture for easy extension
  - http://multicomp.ict.usc.edu/
- hCRF library, machine learning library
  - Matlab and C++ implementations of LDCRF, HCRF and CRF.
  - 2559 downloads during the last year
  - http://hcrf.sf.net/
- GAVAM + CLM-Z, real-time nonverbal behavior recognition
  - Real-time head position and orientation estimation
  - 66 facial feature tracking with automatic initialization
  - http://multicomp.ict.usc.edu/

Thank you!

USC Institute for Creative Technologies