Tracking Continuous Emotional Trends of Participants during Affective Dyadic Interactions using Body Language and Speech Information

Angeliki Metallinou, Athanasios Katsamanis and Shrikanth Narayanan
Signal Analysis and Interpretation Lab (SAIL), USC
contact: metallin@usc.edu
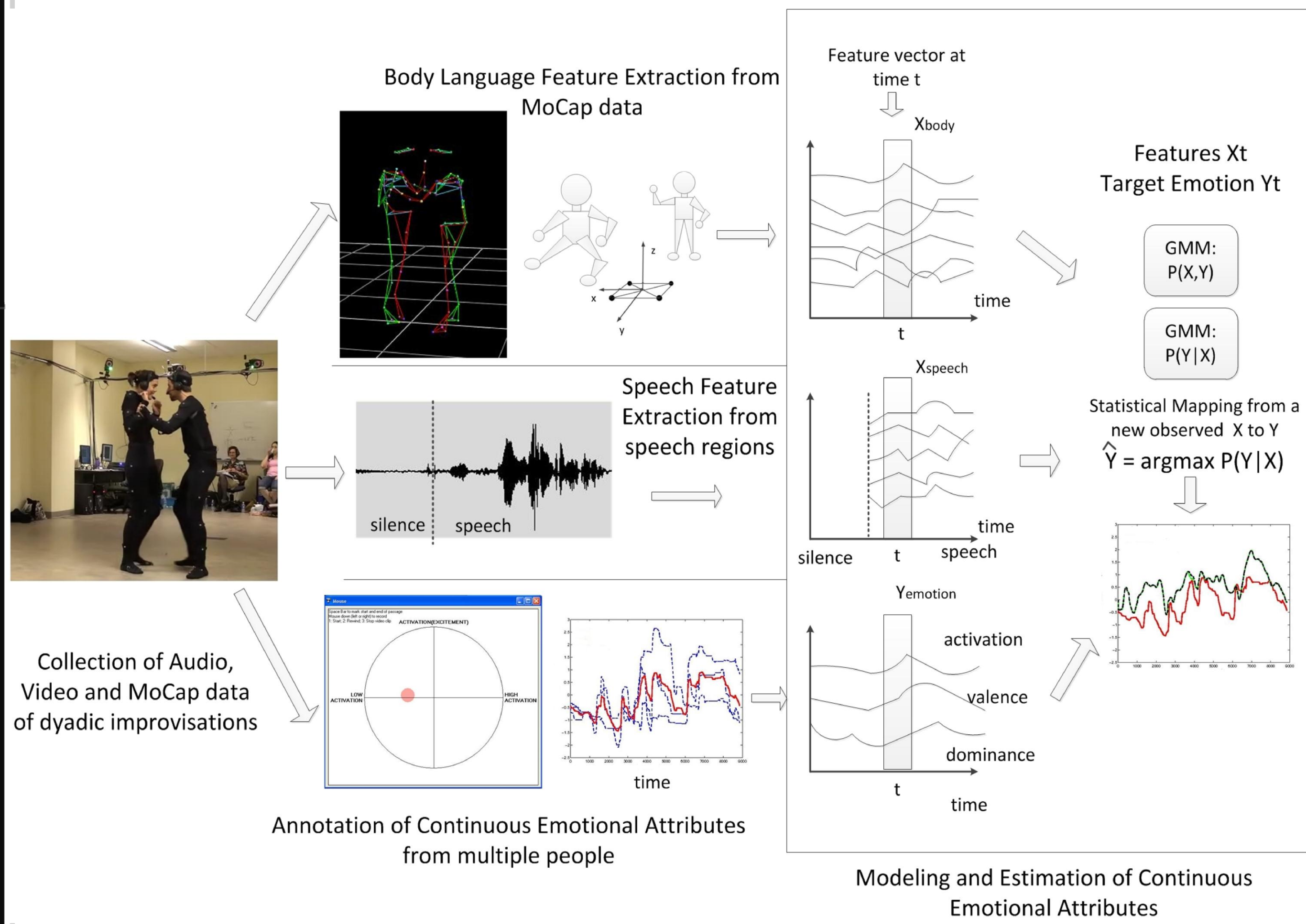
USC Viterbi School of Engineering

SAiL

Goals

- Dynamically track emotional states
  - Activation, dominance
  - Based on body language and speech
- Analysis of body language and emotion
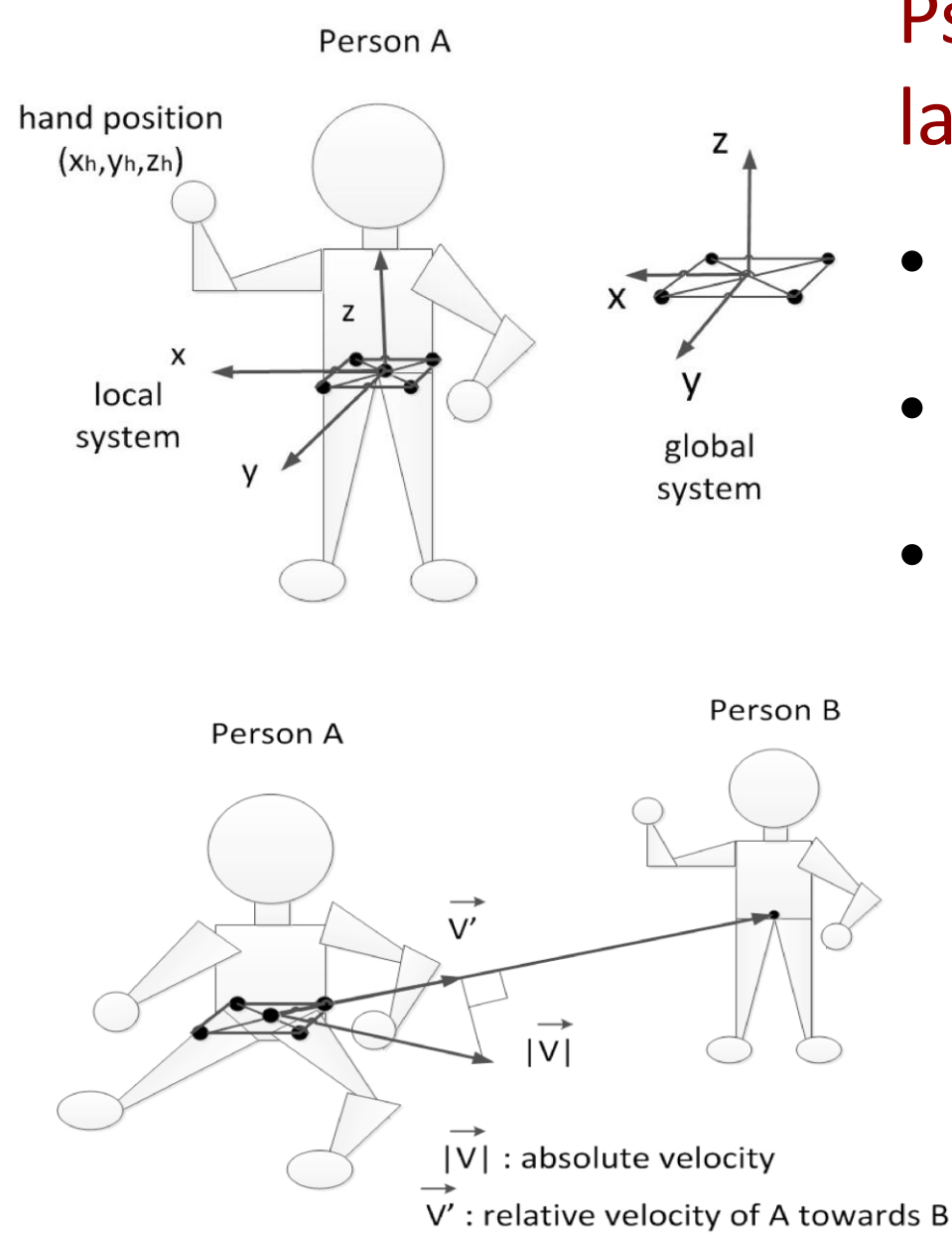- Discuss data labeling/ground truth issues

Findings

- Better at *tracking trends*
- Good at tracking activation trends, modest results for dominance
- Proposed method outperforms regression-based baseline
- Body conveys rich emotional information
  - Hand gestures
  - Body, face orientation
  - Approach/avoidance behaviors

Framework Overview

Body Language Feature Extraction from MoCap data

Speech Feature Extraction from speech regions

silence   speech

Collection of Audio, Video and MoCap data of dyadic improvisations

Annotation of Continuous Emotional Attributes from multiple people

Feature vector at time t

x_{body}

time

x_{speech}

silence   t   speech   time

$y_{emotion}$

activation

valence

dominance

t   time

Features Xt Target Emotion Yt

GMM: $P(X,Y)$

GMM: $P(Y|X)$

Statistical Mapping from a new observed X to Y

$\hat{Y} = \arg\max P(Y|X)$

Modeling and Estimation of Continuous Emotional Attributes

CreativeIT Database and Annotation

- Engineering and Theater
- Theatrical improvisations
- Study expressive body language during interaction
- Continuous annotation
  - Activation, dominance, valence
  - Feeltrace Tool
  - Recruited and trained annotators

Person A
hand position (x_h, y_h, z_h)
local system
global system

Person A   Person B

Person B   Person A

angle

$|V|$: absolute velocity
V' : relative velocity of A towards B

Feature Extraction

Psychology-inspired body language features

- Meaningful behaviors
- Motion Capture
- Feature selection
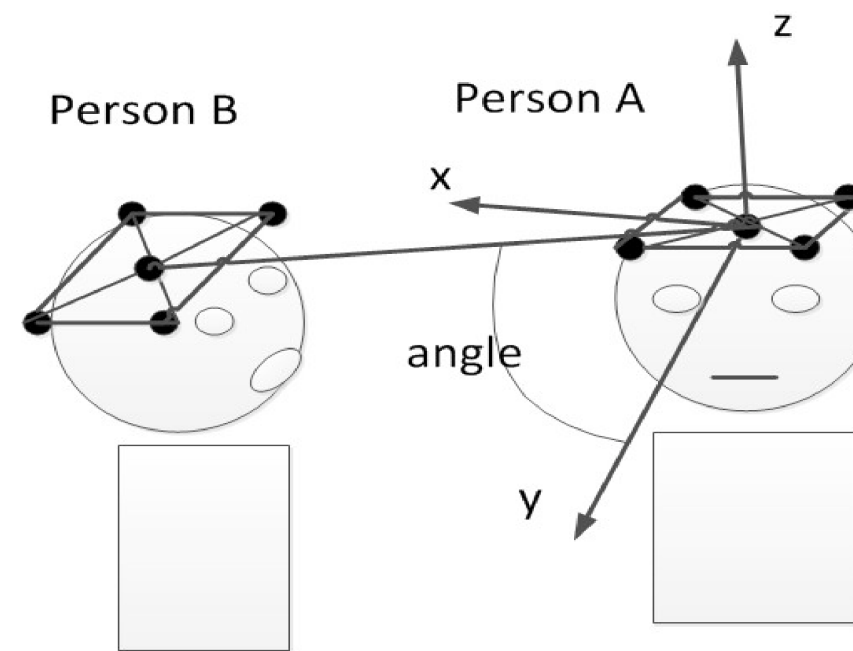
Methodology

Gaussian Mixture Model-based mapping [2]
- Continuous observations (features) y_t
- Hidden emotional states x_t
- Train a joint GMM λ for (x_t, y_t)
- Maximize $P(x_t | y_t, \lambda)$
  - iteratively by EM
- Derivatives for smoother emotional estimates
- Window-level tracking
  - feature functionals
- Visual and audio-visual GMM mappings
- Baseline:
  - Neural Network regression (LSTM)

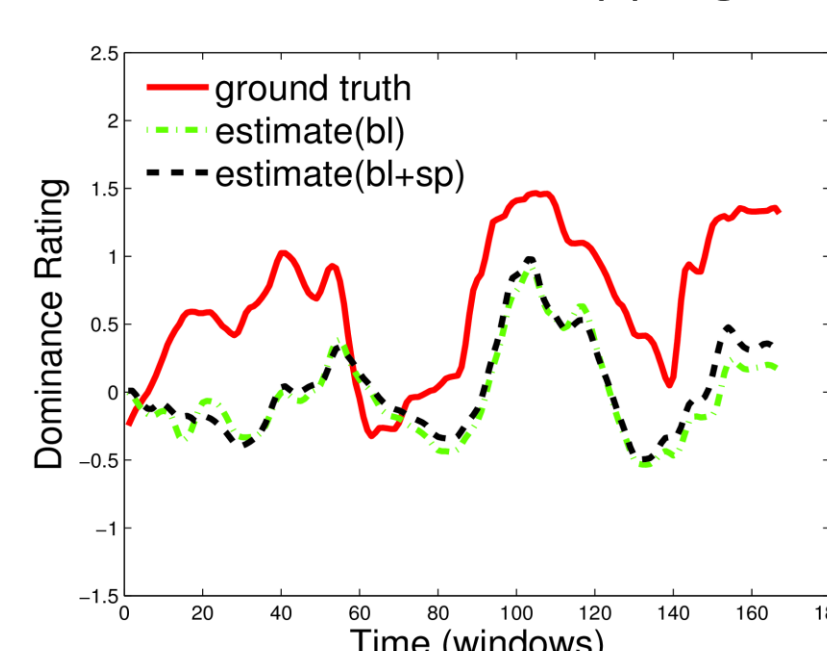## Emotional Body Language Analysis

- High Activation:
  - Higher arm and foot velocities
  - More leaning and orientation towards interlocutor
  - Hands further from body and raised higher
  - Body location at the center of recording space
- High Dominance (relative features are informative):
  - More leaning and orientation towards interlocutor
  - More body/arm/feet moving towards interlocutor
 - More touching

Emotion Tracking Results

Median of correlations between ground truth and emotional estimate

	Body language features		Body language and speech	
	activation	dominance	activation	dominance
GMM-based mapping	0.49*	0.33	0.60*	0.37*
LSTM regression	0.45	0.23	0.49	0.21
Annotator correlations	0.62	0.62	0.62	0.62

GMM-based mapping   LSTM regression

— ground truth
-- estimate(bl)
-- estimate(bl+sp)

Dominance Rating   Time (windows)
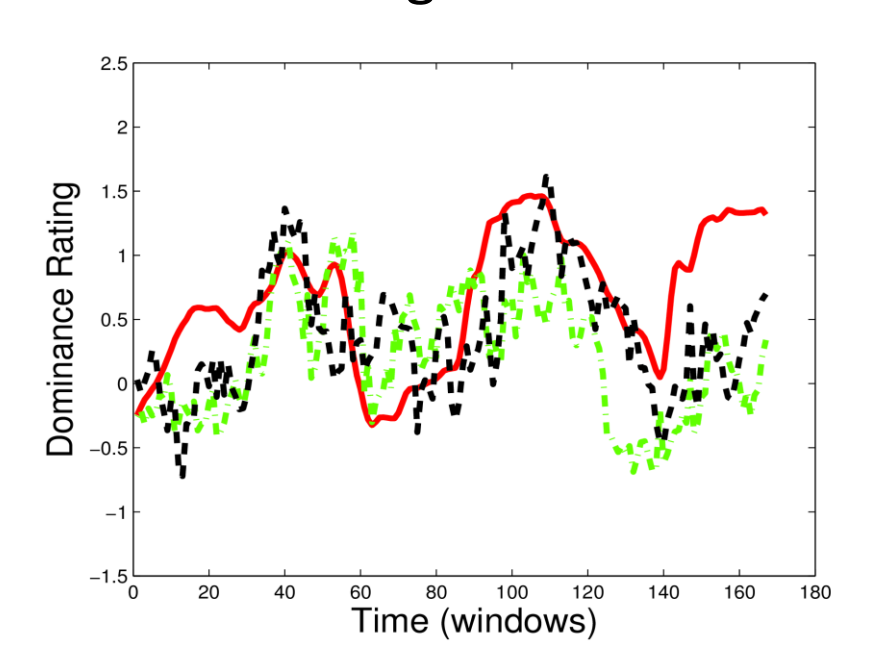
1)A.Metallinou, A. Katsamanis and S. Narayanan, "Tracking continuous emotional trends of participants during affective dyadic interactions using body language and speech information", IMAVIS Special Issue on Continuous Affect Analysis, to appear 2012
2)T. Toda, A. W. Black, and K. Tokuda, "Statistical mapping between articulatory movements and acoustic spectrum using a gaussian mixture model," Speech Communication, 2008