Tracking Changes in Continuous Emotional States using Body Language and Prosodic Cues

Angeliki Metallinou*, Athanassios Katsamanis*, Yun Wang** and Shrikanth Narayanan*

* SAIL lab, EE department, Univ. of Southern California, Los Angeles, CA

** Carnegie Mellon University, Pittsburg, PA

Expressive Interaction

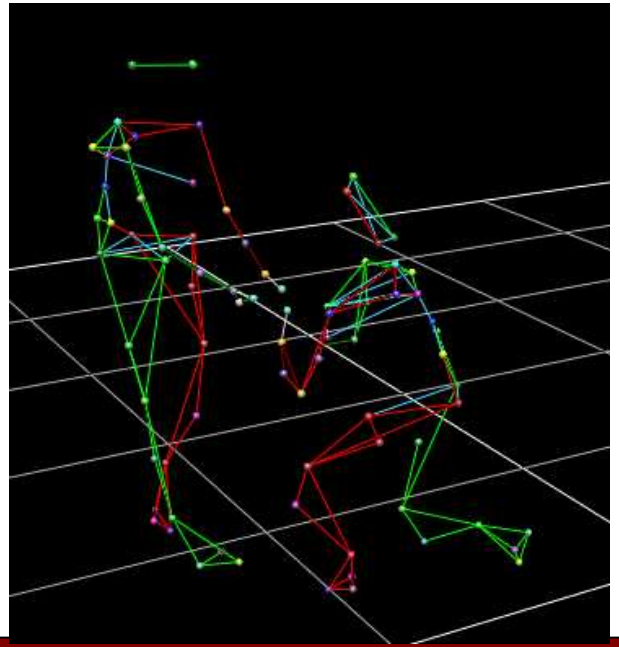
Focus and Goals

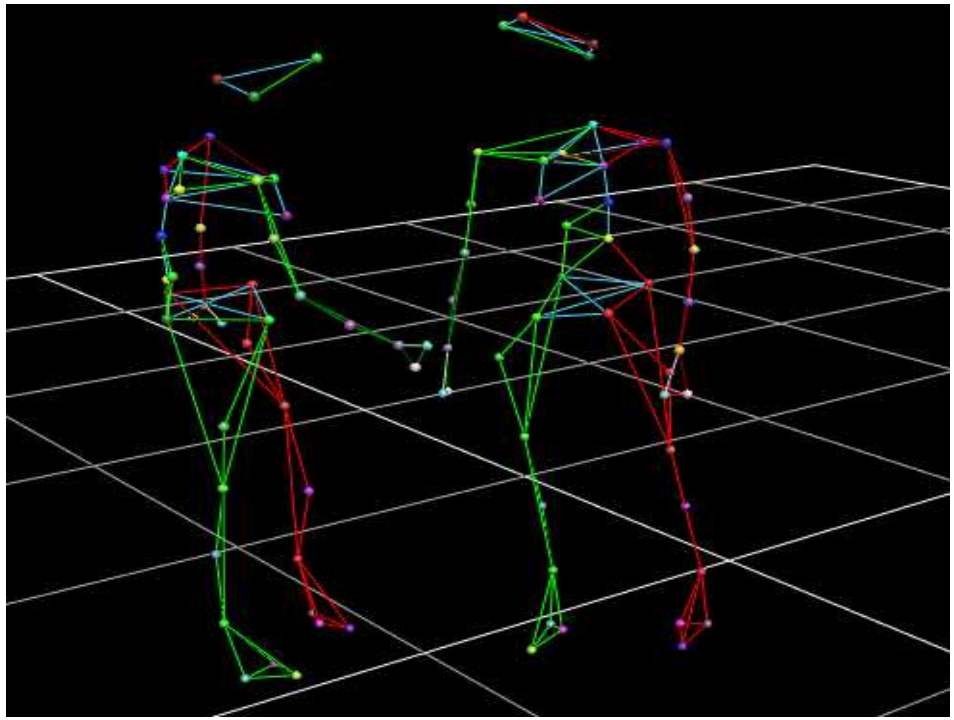
- Examine the **emotional content** of **body language**
  - How are body language cues indicative of the underlying emotional state?

- **Continuously track** unfolding emotional changes
  - Using body language and speech prosody
  - Statistical mapping
    - underlying emotional state
    - observed audio-visual cues

The CreativeIT Database

- The USC CreativeIT database[1]
  - **Multimodal** and **multidisciplinary** database
  - USC Engineering Department and the USC Theatre School
  - Dyadic Theatrical Improvisations
  - Motion Capture, Video, Audio
  - http://sail.usc.edu/improv/

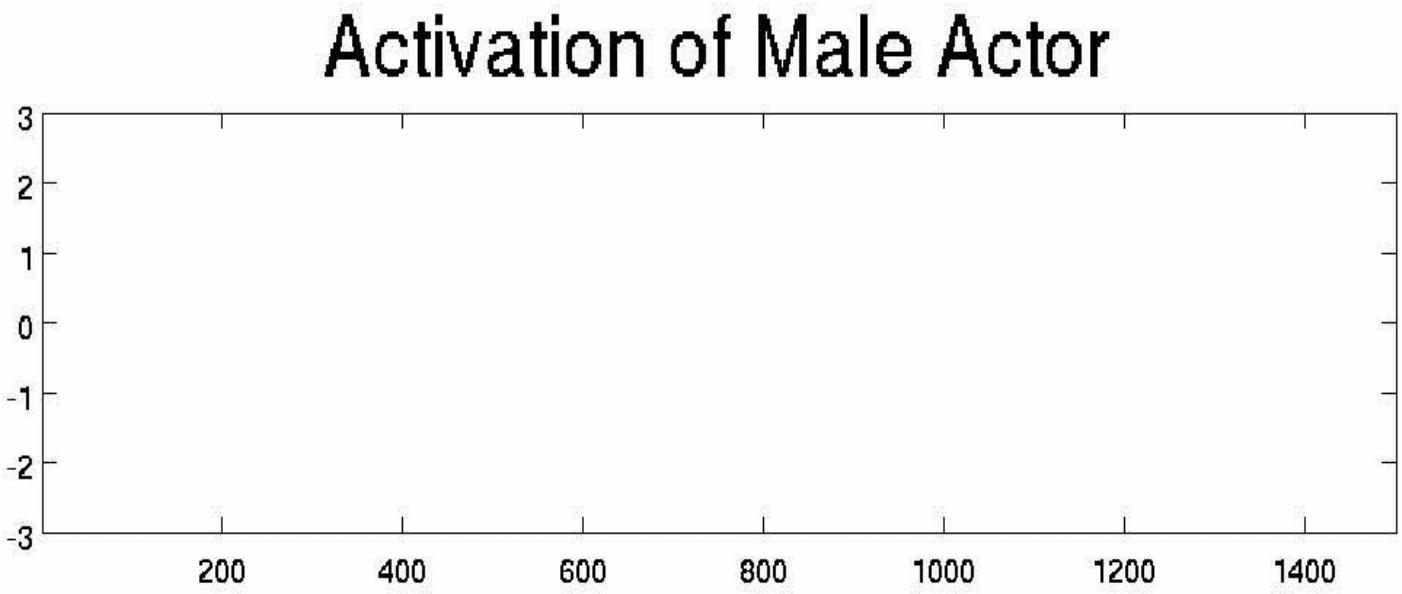




Data Annotation

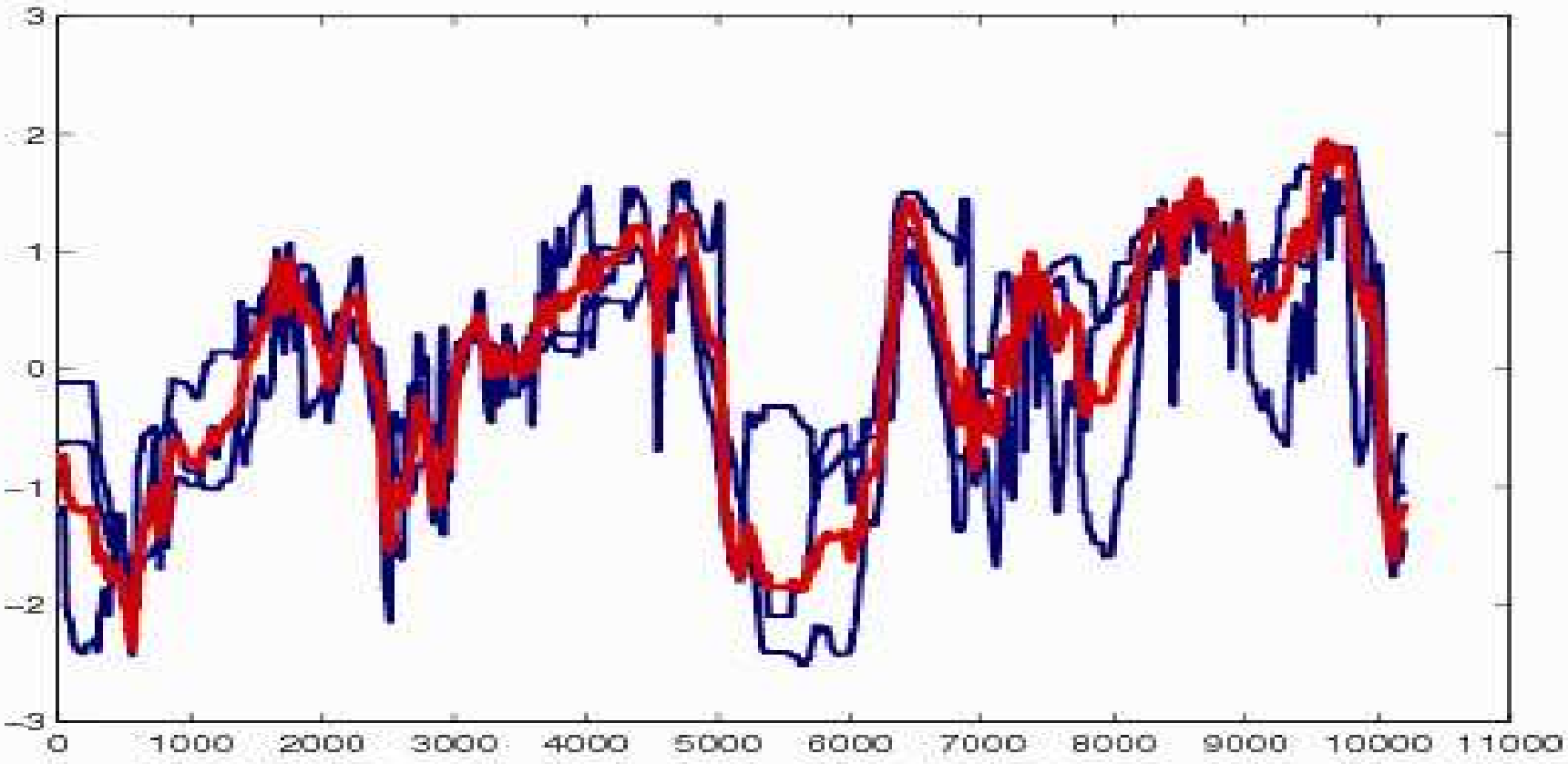
- Emotional Descriptors
  - **Activation**: Excited vs Calm
  - **Valence**: Positive vs Negative
  - **Dominance**: Dominant vs Submissive
- Continuous annotation
  - Using Feeltrace [2]

Example of Annotation from 3 Annotators



Data Annotation

- How to define evaluator agreement?
  - Positive Correlation of annotation curves
  - Average annotator correlation around 0.5



Body Feature Extraction and Analysis

- Body Language Features
  - Intuitive, inspired from psychology
  - **Absolute**
    - Body posture and movement
  - **Relative**
    - Proxemics, touching, looking at, approach/avoidance
    - Convey information about actor interaction
- Statistical Analysis
  - Body language behavior versus
  - Increase, decrease or stability of emotional state

Body Language Features

- Extracted from MoCap markers geometrically
- 20 features in total

Absolute Features	Relative Features
Absolute Velocity of actor A	Angle of A's face towards B
Absolute velocity of A's right hand	Angle of A's body towards B
Absolute velocity of A's left hand	Relative velocity of A towards B
Distance of A's hands	Angle of A's body leaning towards B
Angle of A's head (looking up, down)	Min distance of A's right hand to B's head
….	…..

- From low level features to high level behaviors
  - Relative velocity
  - Moving Towards or Away from other
  - Approach/avoidance

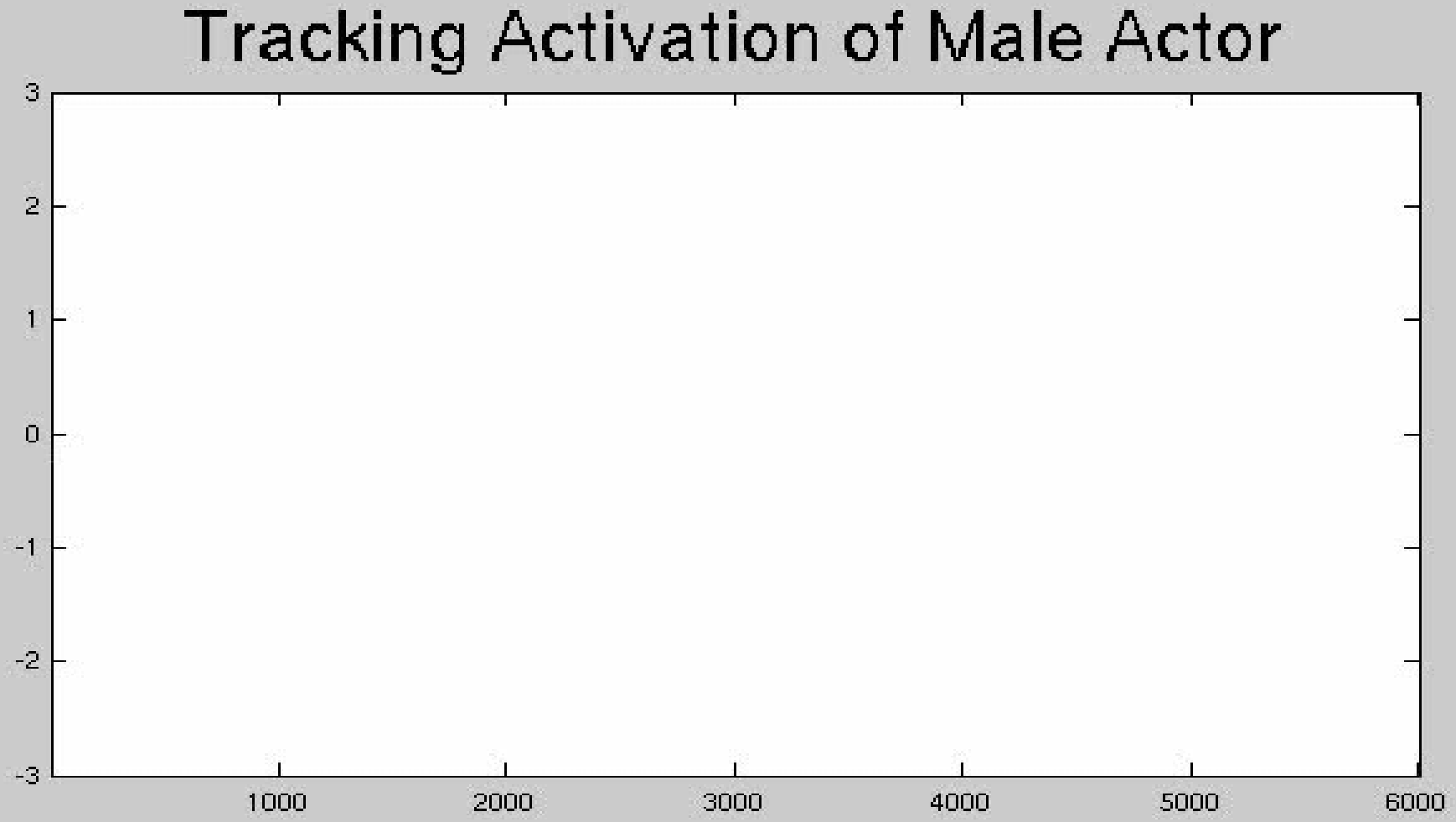
Statistical Analysis of Body Behavior

- Use ground truth to select regions of
  - Increase/decrease/stability of an emotional attribute
- Statistical comparison of body behaviors across regions
- Meaningful results for Activation and Dominance

Activation Increase	Activation Decrease
Walk more	Move away more
Move towards Other more	Hands touch each other more
Move hands towards Other more	

Dominance Increase	Dominance Decrease
Face looking towards Other more	Face looking opposite from Other more
Body leaning towards Other more	Body looking opposite from Other more
	Body leaning away from Other more

Tracking Emotion Changes



Tracking Emotion Changes

- Gaussian Mixture Model-based mapping [3]
  - Continuous underlying emotions $\mathbf{x_t}$
  - Continuous observed body language (and prosody):$\mathbf{y_t}$
  - Train a joint GMM for $\mathbf{(x_t,y_t)}$

$$P(\mathbf{x_t}|\mathbf{y_t},\lambda^{(\mathbf{x},\mathbf{y})}) = \sum_{m=1}^{M} P(m|\mathbf{y_t},\lambda^{(\mathbf{x},\mathbf{y})})P(\mathbf{x_t}|\mathbf{y_t},m,\lambda^{(\mathbf{x},\mathbf{y})}).$$

$$\hat{\mathbf{x}}_t = arg\max_{\mathbf{x_t}} P(\mathbf{x_t}|\mathbf{y_t},\lambda^{(\mathbf{x},\mathbf{y})}).$$

- Iterative process (through EM)
  - Converges to the maximum likelihood mapping (MLE)

- Use derivatives to take into account temporal context
  - **Smoother** emotional trajectory estimates

Using body language and speech

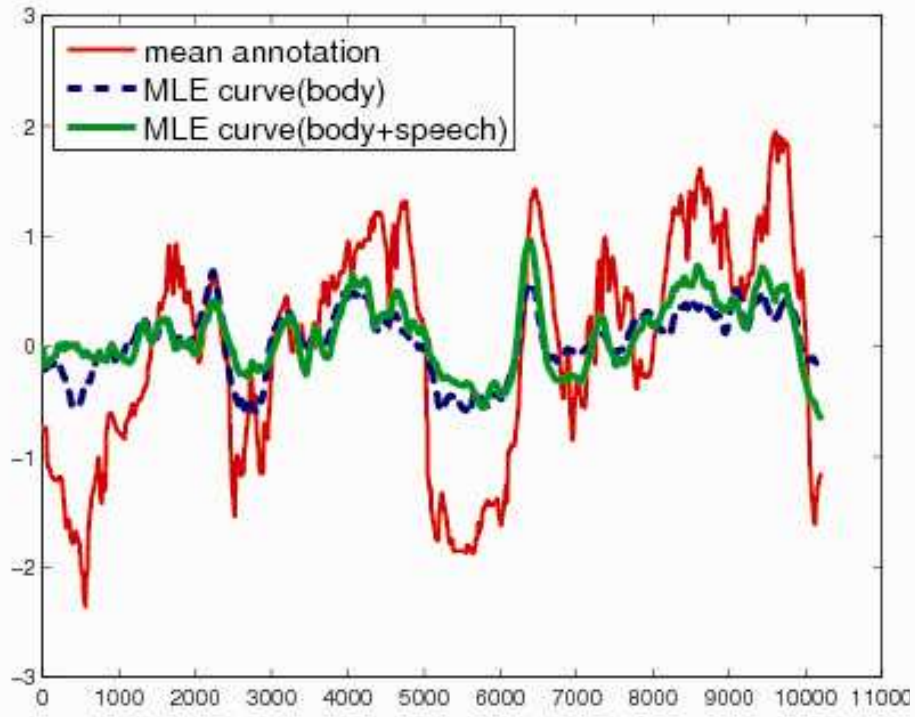
- Body language throughout the interaction
- Audio cues are only relevant when actor is speaking

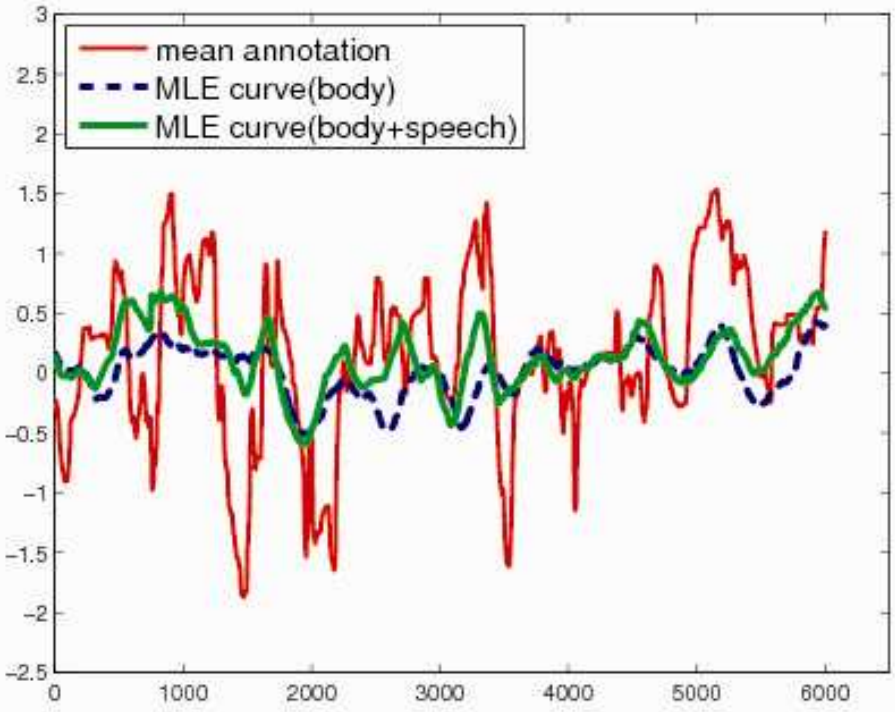
- Train two GMM mappings
  - A **visual GMM** (body features)
  - An **audio-visual GMM** (body and speech features)

- A test recording is chopped into consecutive overlapping segments
  - In each segment the appropriate mapping is applied

- Audio Features
  - Pitch, energy, mel-filterbank features

Some tracking results





- Trajectories follow trends rather than the absolute values
  - Difficulty in quantifying emotions in absolute terms
  - We track emotion changes rather than absolute emotions

Overall results

- Performance of tracking changes
  - **Correlation** between ground truth and estimated curve
  - **Upper bound**: inter-annotator correlations for a recording

Median Correlation	MLE Visual Mapping	MLE audio-visual Mapping	Inter-Annotator Correlations
Activation	0.31	0.42	0.55
Dominance	0.26	0.23	0.47

- For valence we could not track changes
  - Median correlation around zero

Discussion

- Observability of underlying emotional states through our features
  - Capture activation changes,
  - Some of dominance changes
  - But not valence
- Body language and prosody may be more informative of activation and dominance states
- Using prosodic cues greatly benefits activation tracking
- We capture emotional changes rather absolute values of emotional descriptors

Future Work

- Improved features tailored to each emotional attribute
- Improve data annotation process
  - Can we achieve higher inter-evaluator agreement?
- Work towards
  - Continuous monitoring of emotional states
  - Detection of **emotionally salient regions** in an interaction

References

1) A. Metallinou, C.-C. Lee, C. Busso, S. Carnicke, and S. Narayanan, “The USC CreativeIT database: A multimodal database of theatrical improvisation,” in LREC Workshop on Multimodal Corpora, Malta, 2010.

2) R. Cowie, E. Douglas-Cowie, S. Savvidou, E. McMahon, M. Sawey, and M. Schroeder, “Feeltrace: An instrument for recording perceived emotion in real time,” 2000.

3) Tomoki Toda, Alan W. Black, and Keiichi Tokuda, “Statistical mapping between articulatory movements and acoustic spectrum using a gaussian mixture model,” Speech Communication, vol. 50, pp. 215–227, 2008.

Thank you!

Questions?