EVALUATING TEMPORAL PATTERNS IN APPLIED INFANT AFFECT RECOGNITION

Allen Chang, [1], Lauren Klein*, [1], Marcelo R. Rosales[2], Weiyang Deng[2], Beth A. Smith[3], Maja J. Matarić[1]*

[1] Department of Computer Science, University of Southern California, Los Angeles, USA
[2] Division of Biokinesiology and Physical Therapy, University of Southern California, Los Angeles, USA
[3] Infant Neuromotor Control Lab, Childrens Hospital Los Angeles, Los Angeles, USA

** Equal Contribution*

This project suggests guidelines for evaluating affect recognition performance over time.

We hope you can come away with some guidelines for implementing continuous affect recognition in the presence of:

1. Missing data
2. Ambiguous human behavior

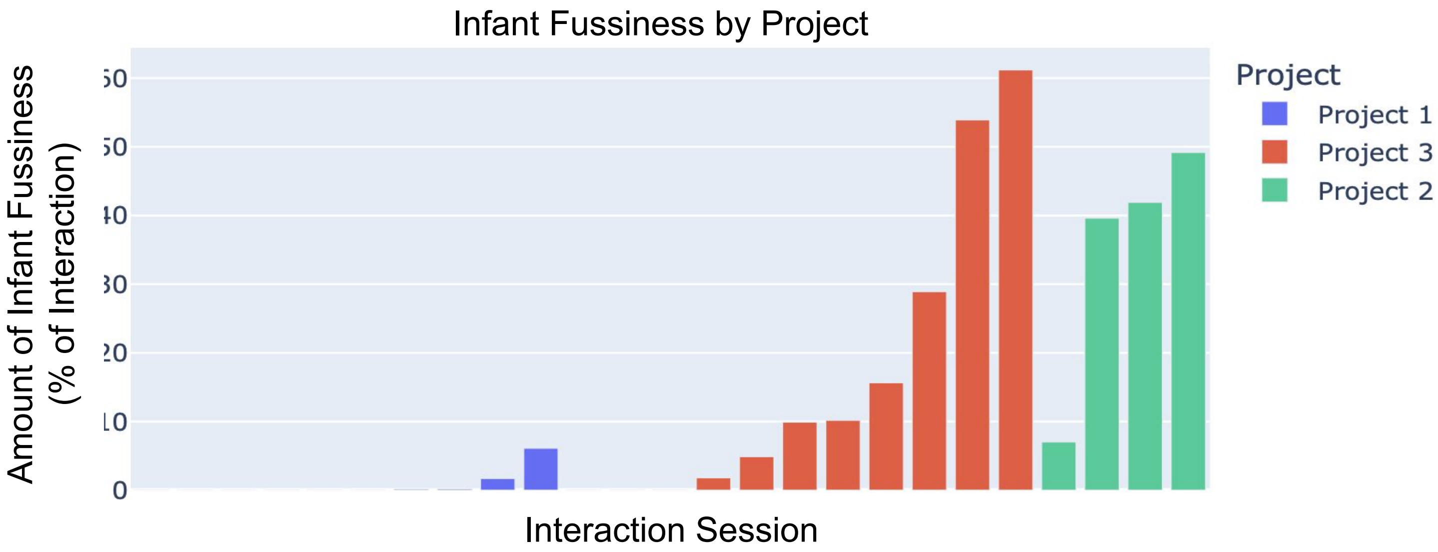
Our lab studied infant-robot interaction with socially assistive robotics (SAR) to promote early development.

Observation:

The infant's affective state changes with the difficulty of the interaction.

!

As interactions became more challenging, infants became more fussy.



Continuous affect recognition helps robots respond to human partners in real-time.

!

Times when continuous infant affect is important:

- **Choosing robot actions**
  - Robots should respond rapidly and adjust actions to respond to human partners
  - Example: Infant-robot interaction
- **Interaction analysis**
  - Affect is used to evaluate synchrony between dyadic interaction partners
  - Example: Assessment of child social interactions

Past work with infant emotion recognition excludes data when faces are occluded.

Messinger, Daniel S., Mohammad H. Mahoor, Sy-Miin Chow, and Jeffrey F. Cohn. "Automated measurement of facial expression in infant–mother interaction: A pilot study." Infancy 14, no. 3 (2009): 285-305.

Lysenko, Sofiya, Nidhi Seethapathi, Laura Prosser, Konrad Kording, and Michelle J. Johnson. "Towards Automated Emotion Classification of Atypically and Typically Developing Infants." In 2020 8th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob), pp. 503-508. IEEE, 2020.

Observation:

Past work typically do not report performance metrics during times where data were missing or during affect transitions.

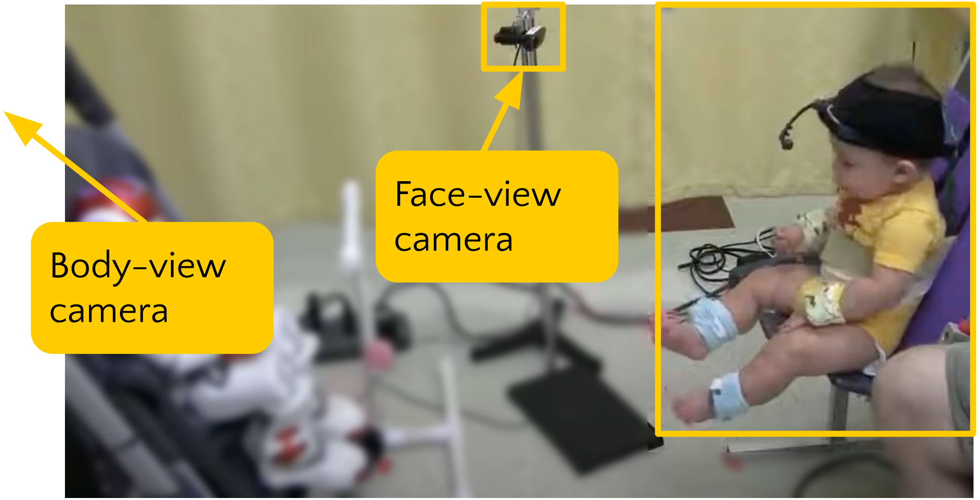
!

Research Questions:

1. Can infant affect recognition be performed with body landmarks (because faces were more often occluded)?
2. Does infant affect recognition performance change for windows of missing data or over time within infant-robot interactions?

The infant-robot interaction was designed to promote exploratory leg movements.

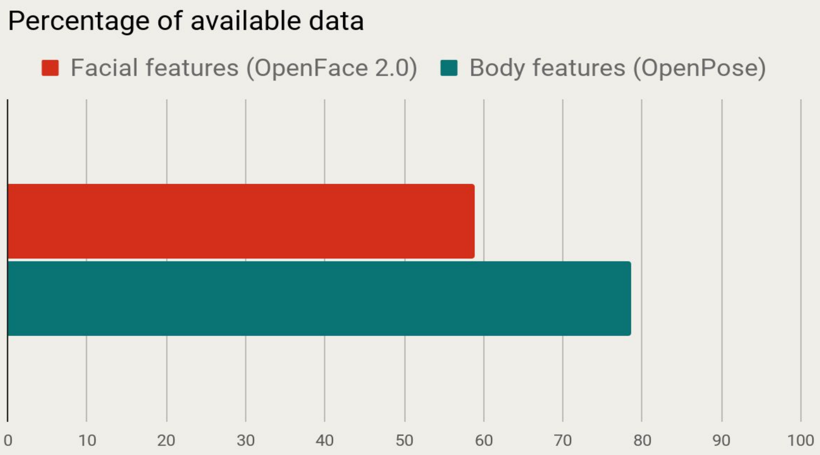
Data was captured as face-view and body-view. Each infant video frame was labeled as "alert" or "fussy."



We extracted facial landmarks, facial action units, and body landmarks.

Data with a majority of features recognized with ≥ 20% confidence for more than 90% of a time window were categorized as $\mathcal{D}_{\text{confident}}$.

All data, regardless of feature extraction confidence, were included in $\mathcal{D}_{\text{total}}$.

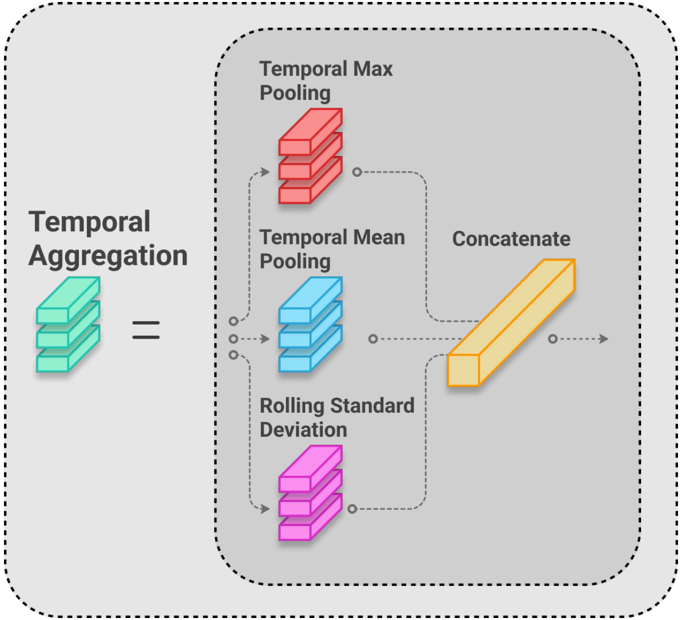


We normalized by distance and aggregated features temporally.

Landmarks were centered and afterwards were scaled with respect to the infant's size.

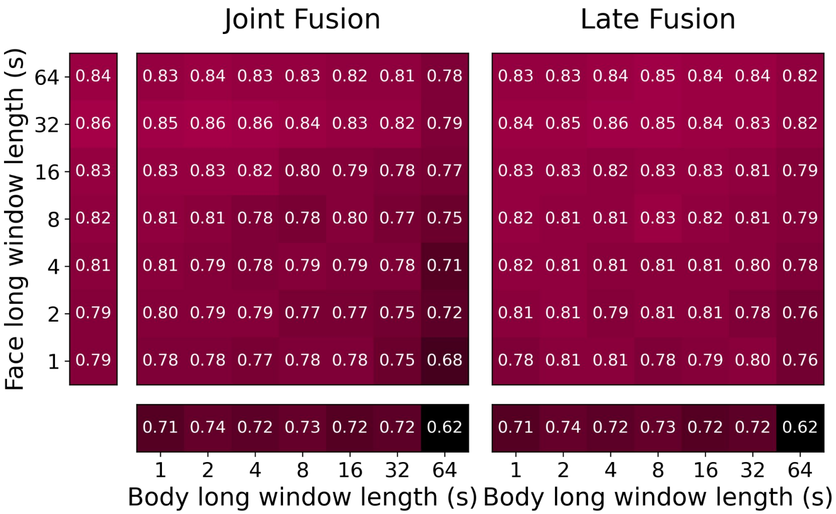
We aggregated features over short and long time windows to capture short- and long-term behavior:

- 0.5 seconds for the short window length
- We tested multiple long window lengths



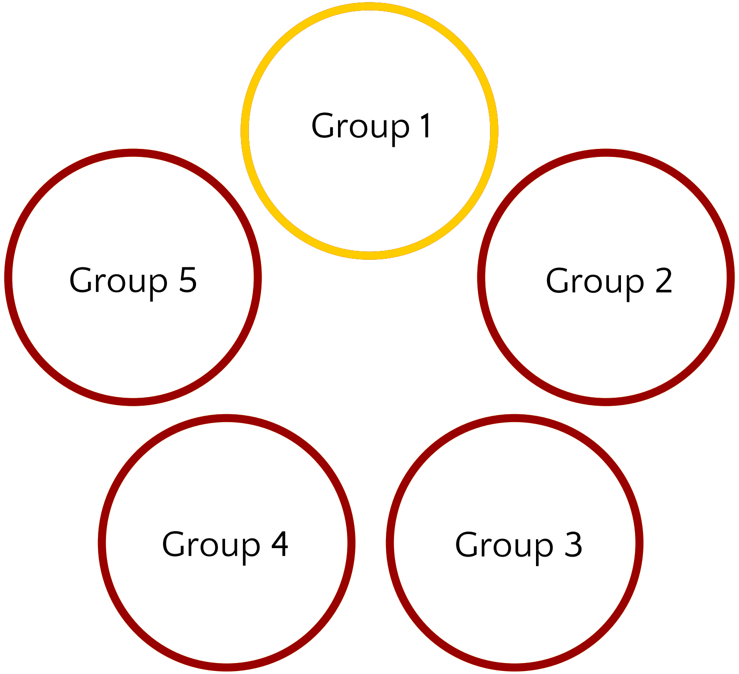
Aggregation methods included in temporal aggregation.

Longer window lengths were optimal for facial features compared to body features.



Mean AUC across stratified cross-validation folds for each set of long window lengths for facial and body features.

We performed cross-validation for 5 groups of infants.

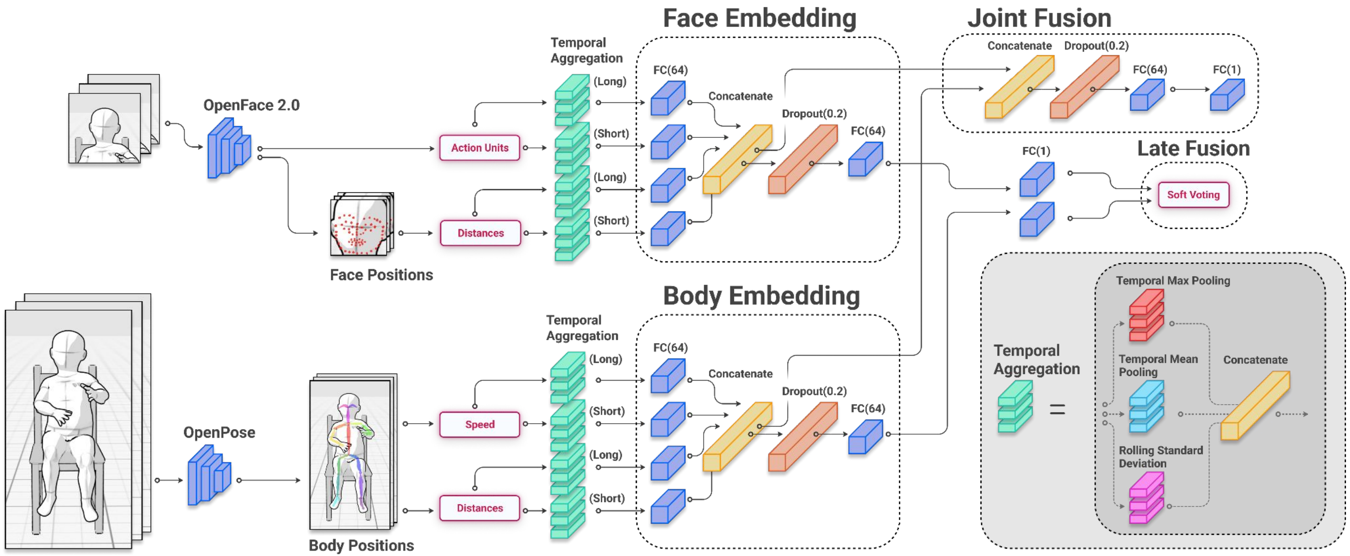


Infant groupings for our affect classification experiments.

Groups were balanced by the amount of fussy data

- Trained on 4 Groups
- Tested on 1 Group

We explored unimodal face models and body models as well as multimodal joint fusion and late fusion models.



Overview of the modeling framework. FC: Fully Connected (Dense) Layer.

We evaluated model accuracy with respect to:

1. **Time since an actual affective state transition**
   (change in recall over time)
2. **Time since a predicted affective state transition**
   (change in precision over time)

On $\mathcal{D}_{\text{confident}}$, the best-performing unimodal and multimodal models reached the same optimal AUC.

Dataset \ Model	Face	Body	Joint Fusion	Late Fusion
$\mathcal{D}_{\text{confident}}$	**0.86**	**0.74**	**0.86**	**0.86**
$\mathcal{D}_{\text{total}}$	0.67	0.65	0.73	0.69

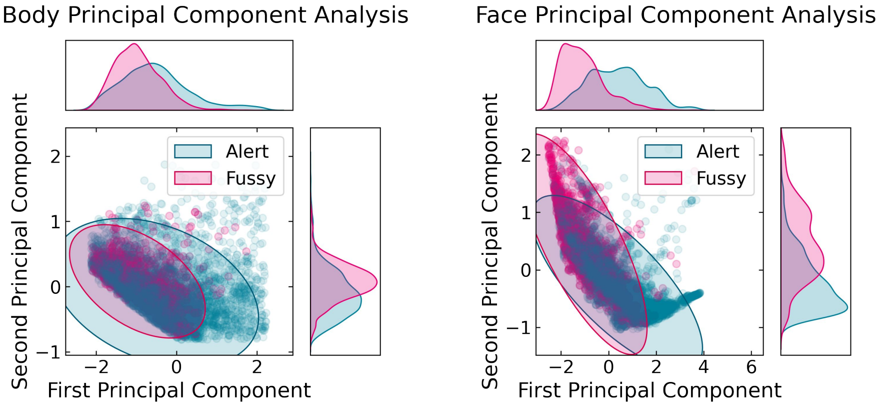
AUC values for best-performing face, body, joint-fusion, and late-fusion models.

On $\mathcal{D}_{\text{total}}$, both multimodal models outperformed each unimodal model.

Dataset \ Model	Face	Body	Joint Fusion	Late Fusion
$\mathcal{D}_{\text{confident}}$	0.86	0.74	0.86	0.86
$\mathcal{D}_{\text{total}}$	**0.67**	**0.65**	**0.73**	**0.69**

AUC values for best-performing face, body, joint-fusion, and late-fusion models.

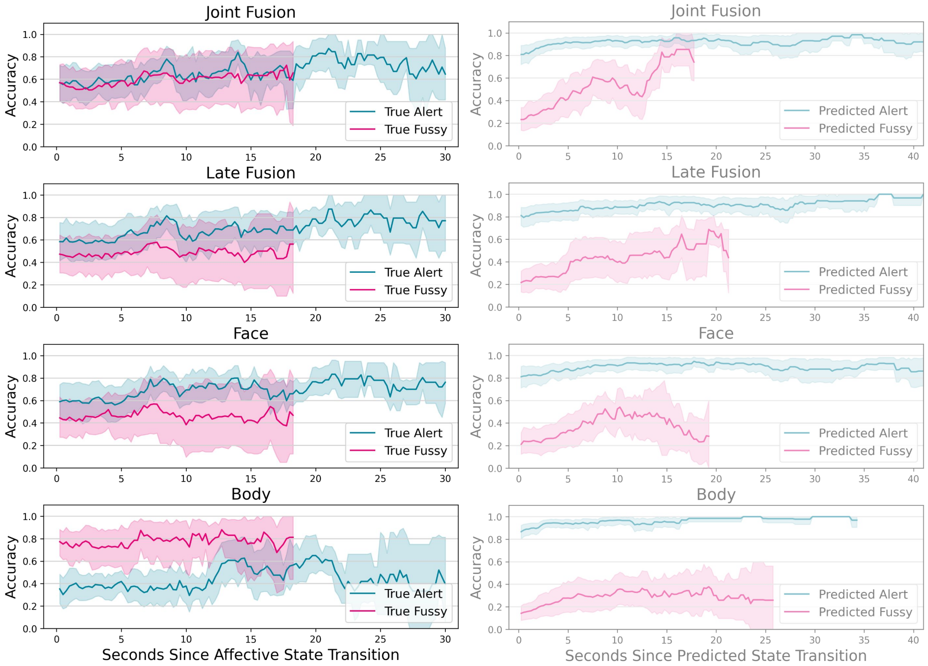
Our results reveal that body features are a viable modality for predicting infant affect.



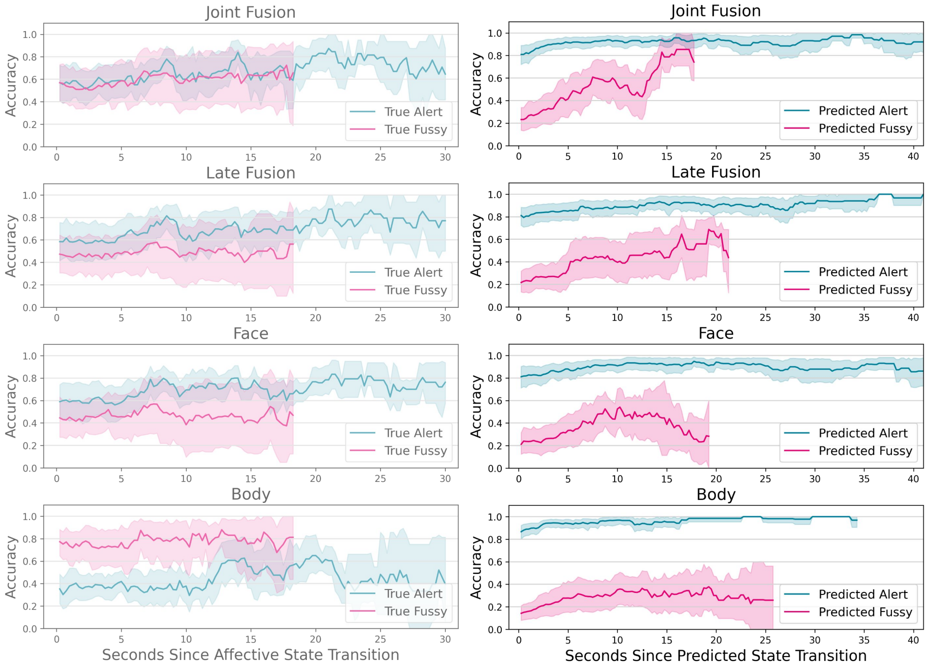
Scatter plot and marginal density distributions of the first two principal components of face and body embeddings, for 1 test fold of infants. Ellipses contain data points within 3 standard deviations of the mean.

Model recall for “fussy” → “alert” generally increased as the time since an actual affective state transition increased.

Model accuracy versus time since an **actual** affective state transitions.



Model precision for "alert" → "fussy" increased with the time since a predicted affective state transition.



Model accuracy versus time since a **predicted** affective state transitions.

Key Takeaways:

- Body features can be used for infant affect recognition.
- Multimodal models outperformed unimodal models when evaluating on data with missing features.
- Model precision and recall varied when data were missing and over time.

Thank you!

This work was supported by the National Science Foundation (NSF CBET-1706964).

Infant Neuromotor Control Laboratory

Key Takeaways:

- Body features can be used for infant affect recognition.
- Multimodal models outperformed unimodal models when evaluating on data with missing features.
- Model precision and recall varied when data were missing and over time.