# DYNAMIC MULTI-RATER GAUSSIAN MIXTURE REGRESSION INCORPORATING TEMPORAL DEPENDENCIES OF EMOTION UNCERTAINTY USING KALMAN FILTERS

Never Stand Still

School of Electrical Engineering & Telecommunications

*Ting Dang*[1,2], *Vidhyasaharan Sethu*[1], *Eliathamby Ambikairajah*[1,2]

[1] School of Electrical Engineering and Telecommunications, UNSW, Australia
[2] ATP Research Laboratory, DATA61 (CSIRO), Australia

# Content

1. **Continuous Emotion Prediction**
2. **Inter-rater Variability**
3. **Dynamic multi-rater GMR**
4. **Experimental Results**
5. **Conclusion**

# Continuous Emotion Prediction

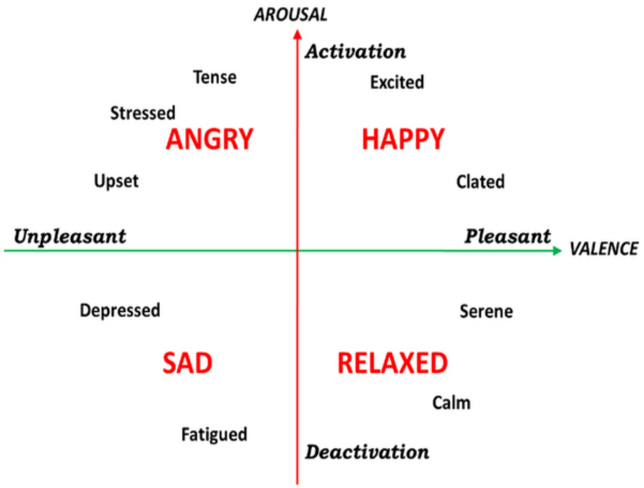
## ➢ Emotion Representation

- **Categorical Representation**

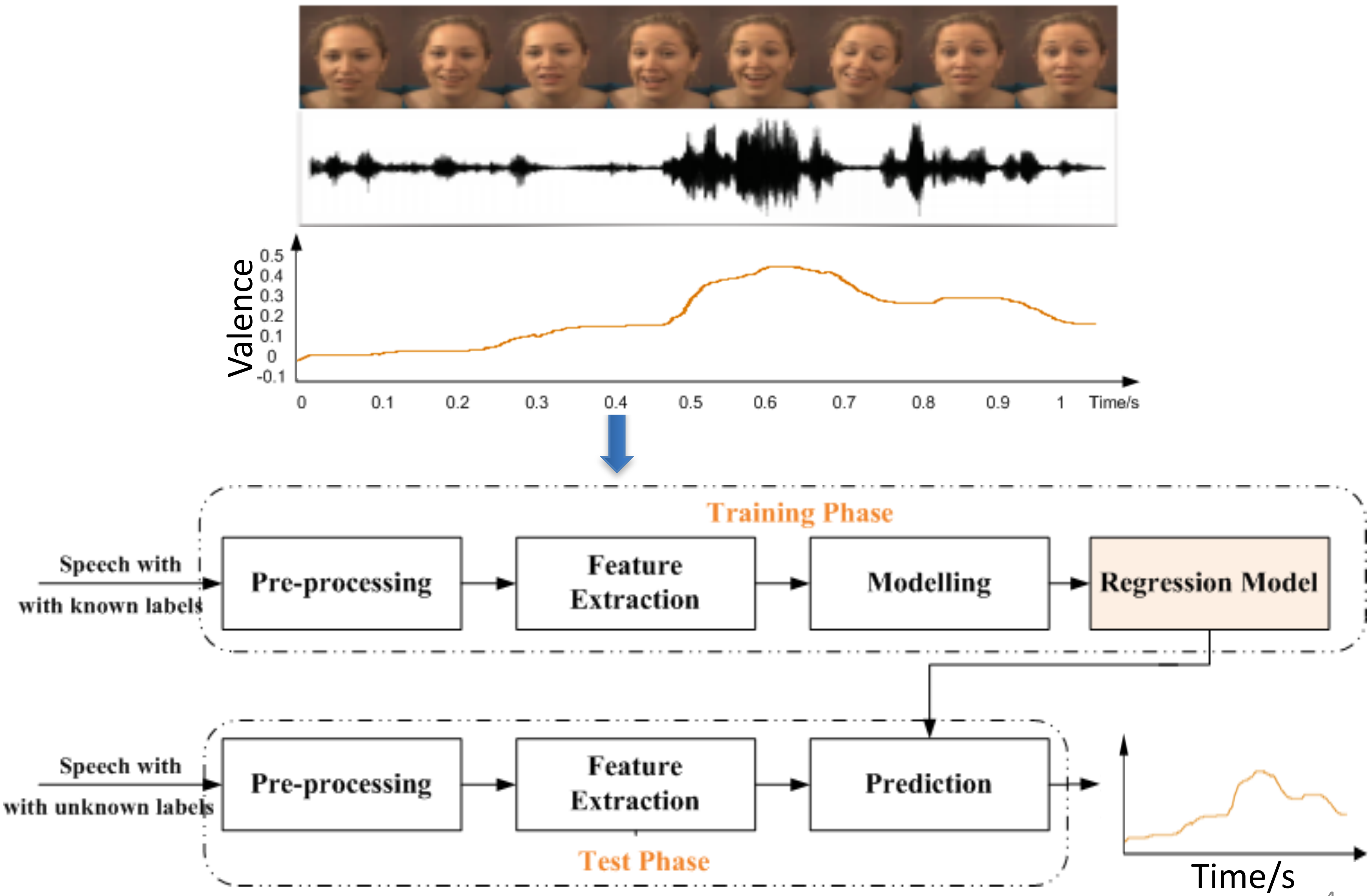
--- Happy, anger, sad, etc.


- **Dimensional Representation**

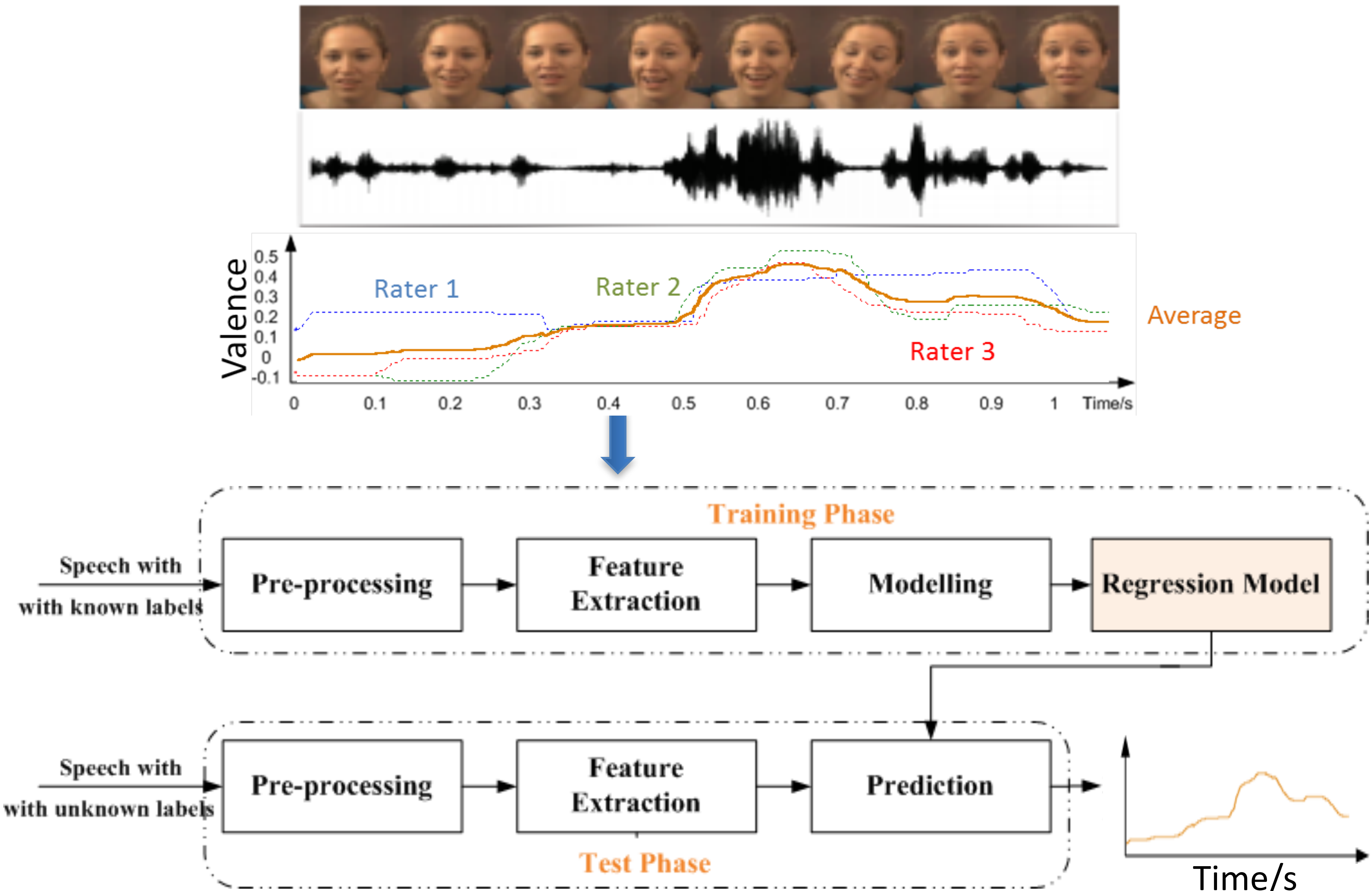
--- Affective attribute: arousal, valence



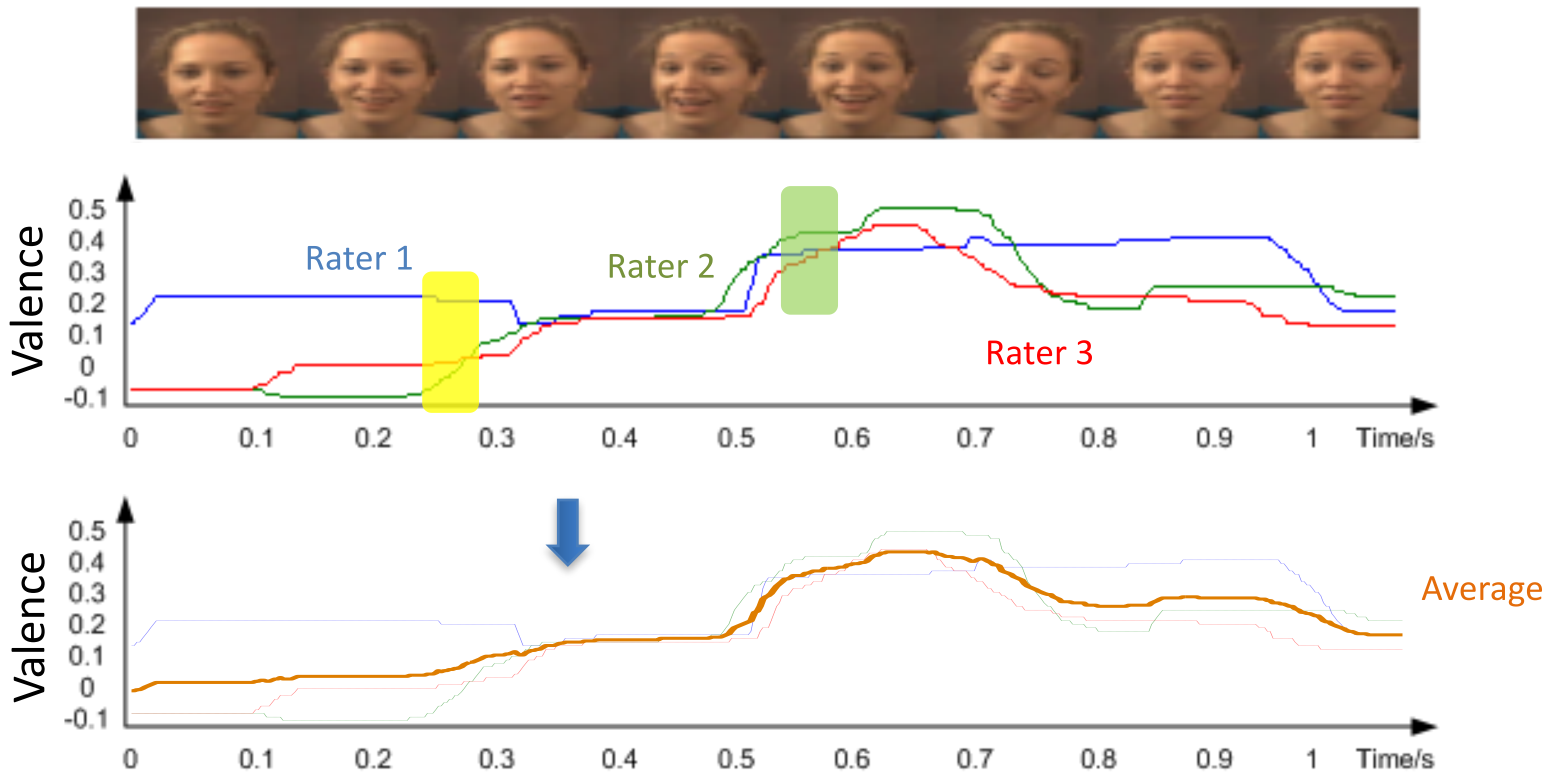
# Continuous Emotion Prediction



# Continuous Emotion Prediction

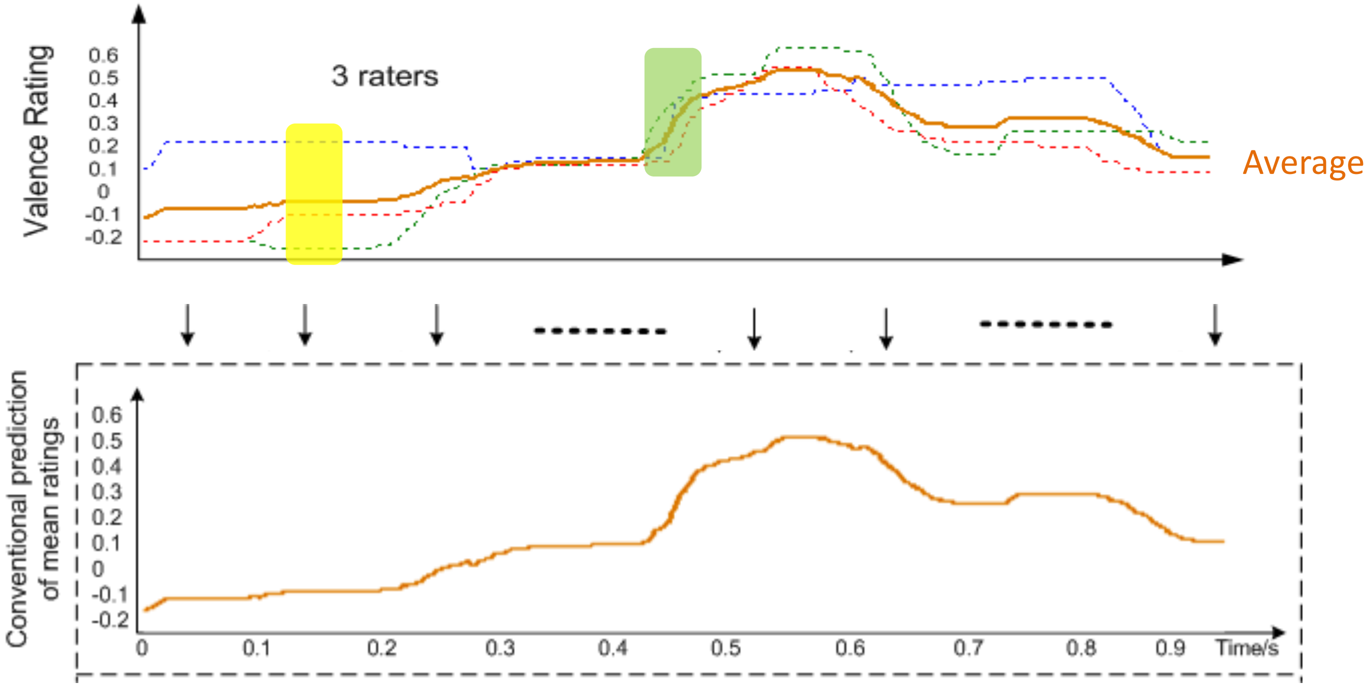


# Inter-rater Variability

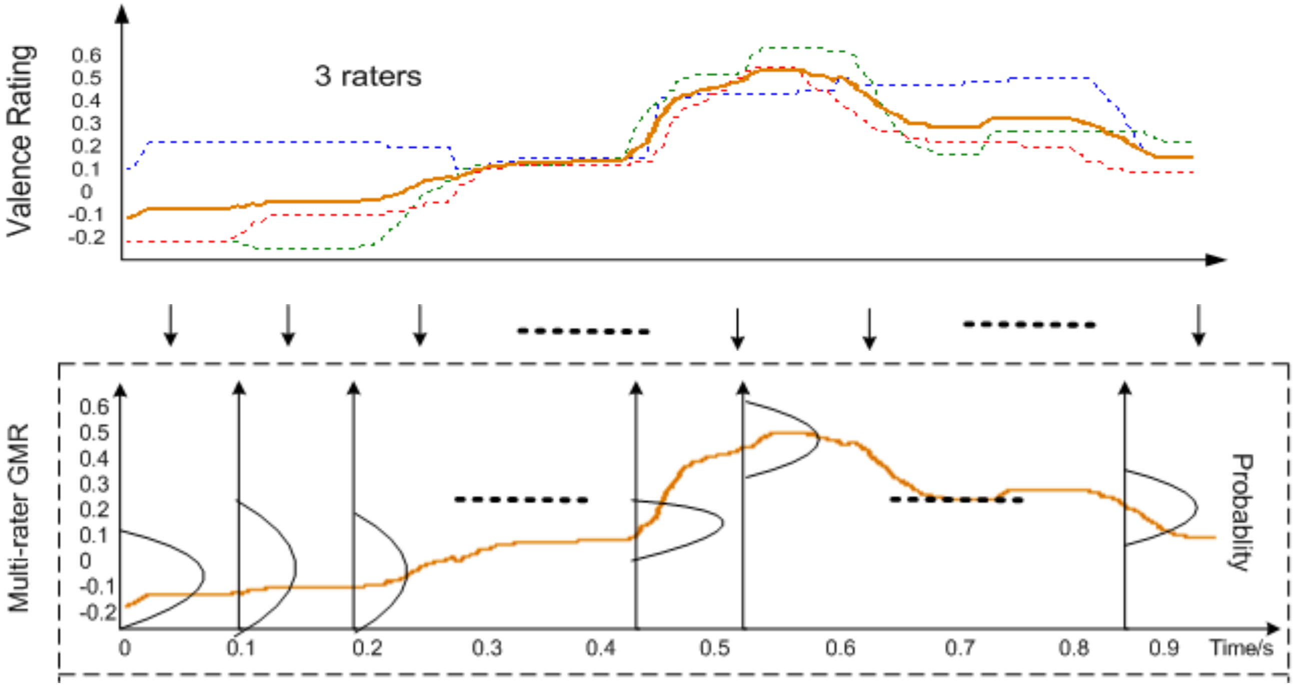



- Averaging ratings ignore the discrepancies between raters
- Intense emotions are easier to recognize while the subtle emotions are more ambiguous.
- Other factors (i.e. recording conditions) may affect rater’s judgements

# Inter-rater Variability

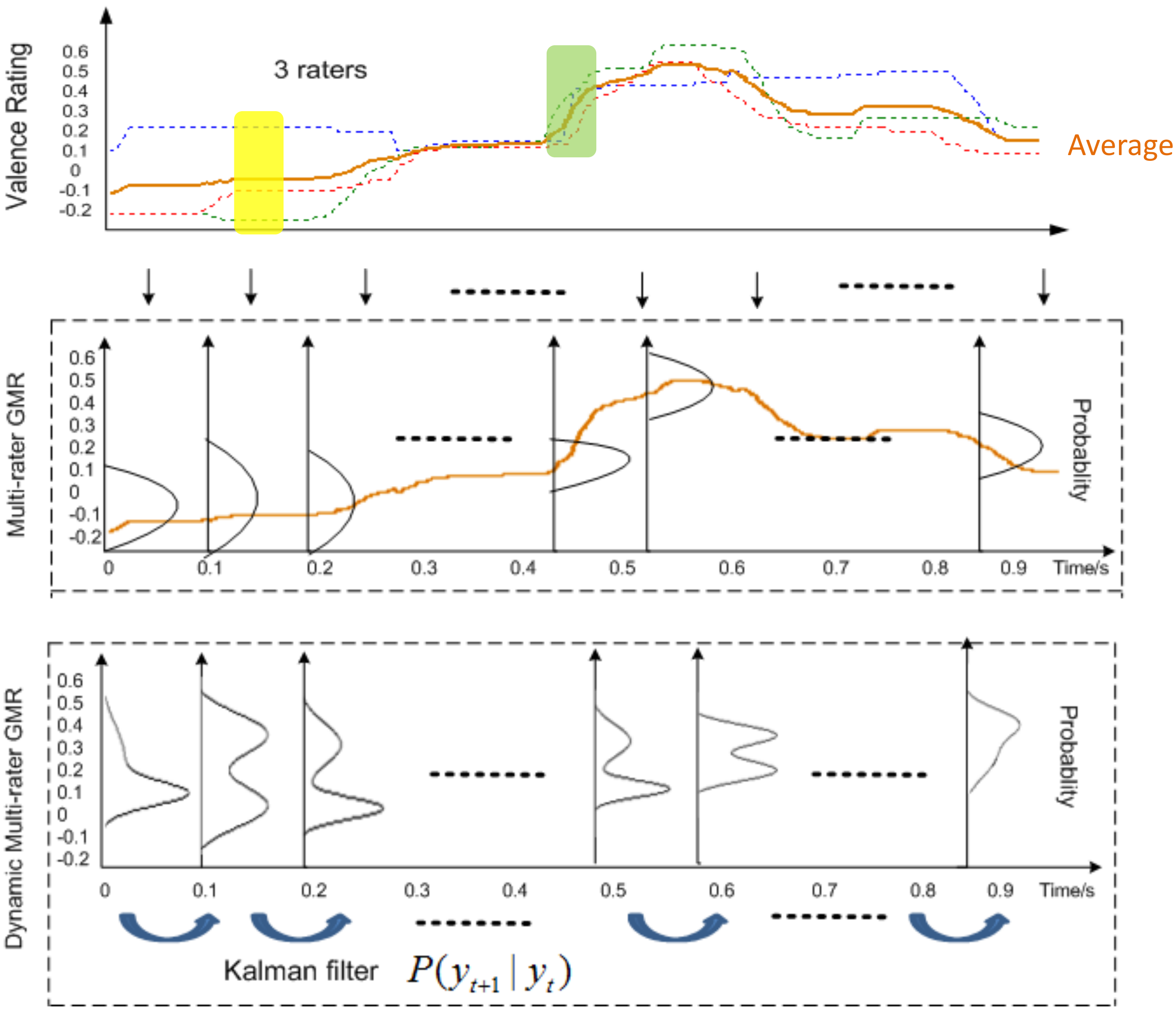


# Inter-rater Variability

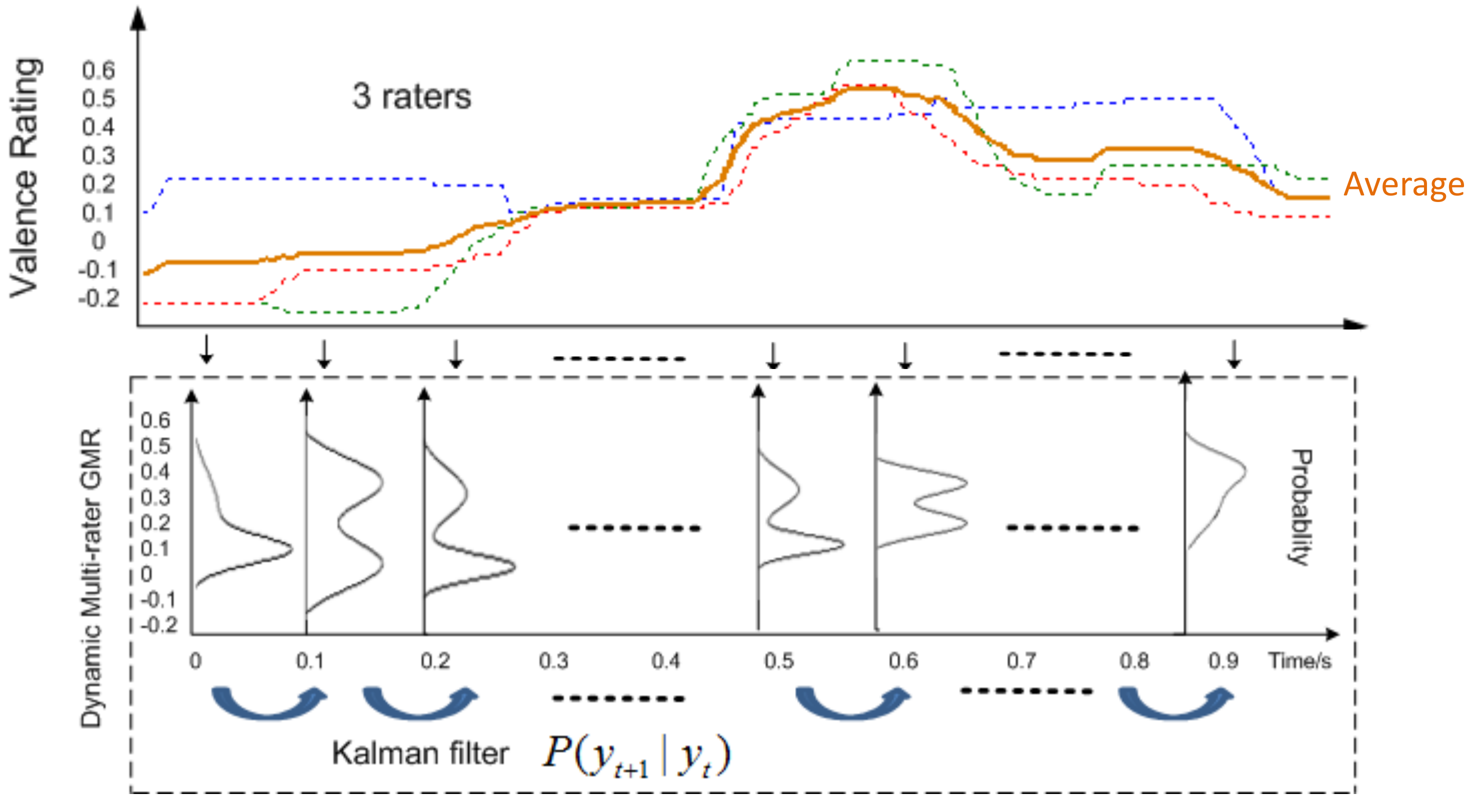



- Gaussian assumption of label distribution may not be true
- Multi-rater Gaussian mixture regression (GMR) does not consider temporal dependencies

# Inter-rater Variability



# Dynamic multi-rater GMR

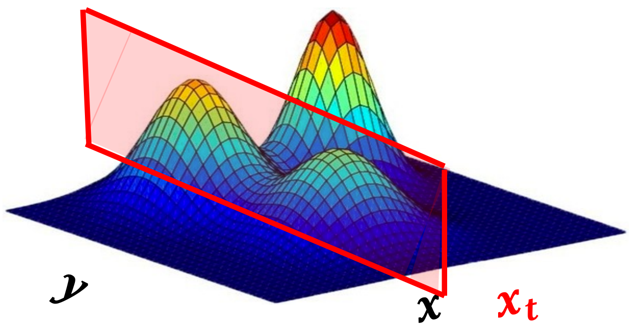



- Incorporation of both *forward and backward Kalman filters* into multi-rater GMR to account for the *temporal dependencies in both directions*.
- *Label distribution given by GMM* instead of single Gaussian.
- Measure to *quantify uncertainty* from predicted distribution (GMM).

# Gaussian Mixture Regression(GMR)

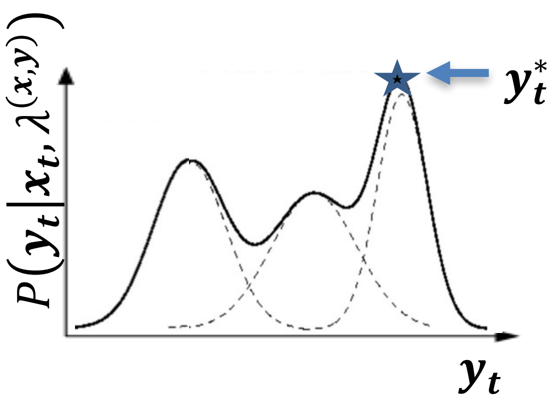
## GMR model

- Joint distribution of feature vectors and labels

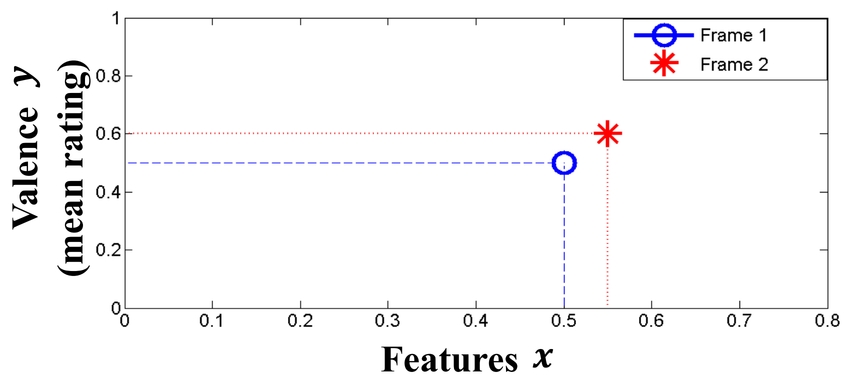



$\lambda^{(z)} = P(x, y)$

## Probability distribution

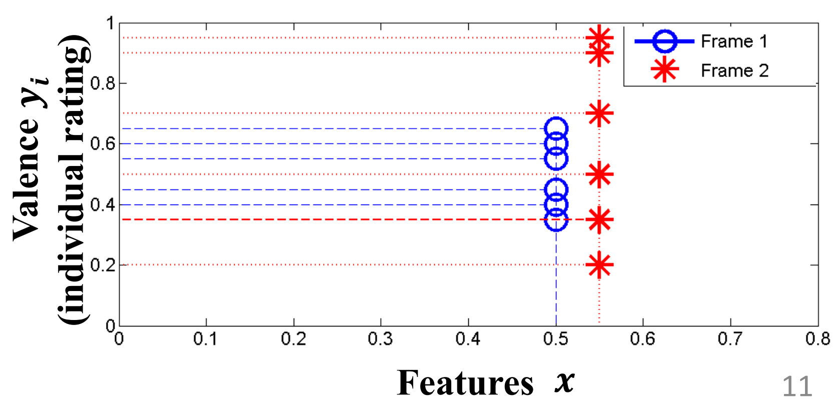



- Training vectors are generated by concatenating the feature vector and *mean rating*

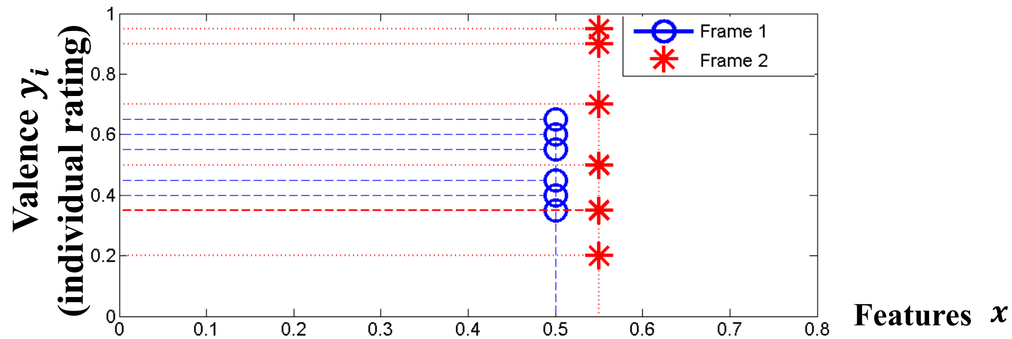



## Incorporation of uncertainty

- Training vectors are generated by concatenating the feature vector and *individual annotation*

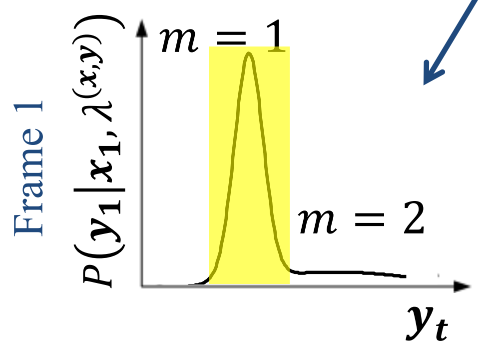


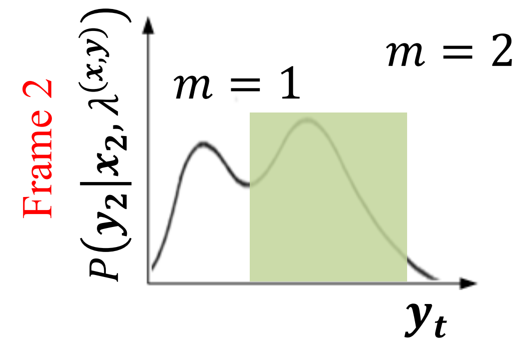
# Gaussian Mixture Regression(GMR)




## ➢ Predicted label distribution

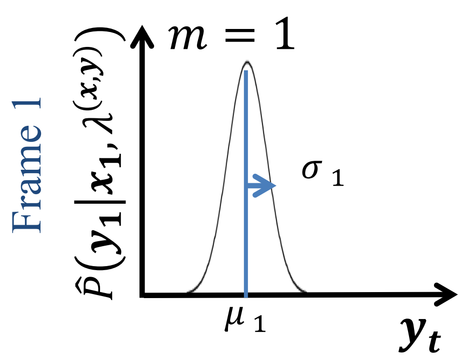
- Predicted label distribution (GMM)

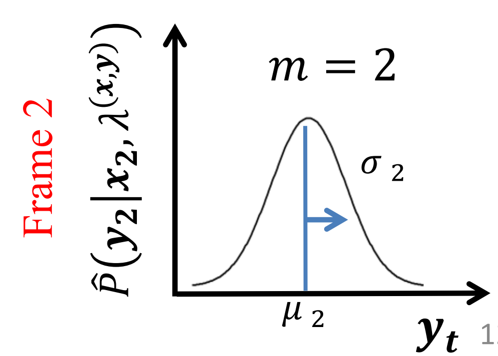



Dominant mixture component to approach the label distribution

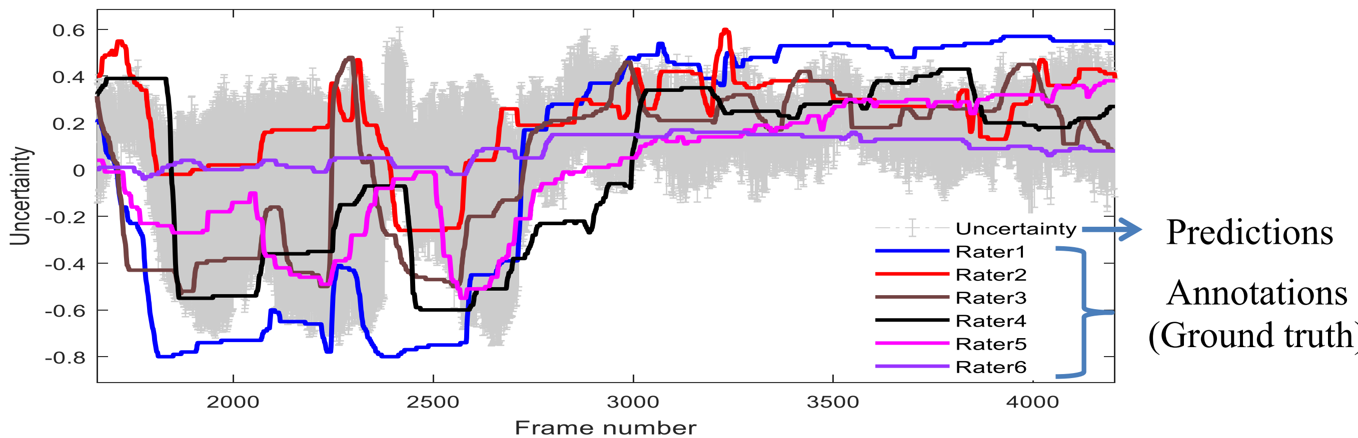
- Approximated label distribution (Gaussian)



# Gaussian Mixture Regression(GMR)

## ➢ Plot of uncertainty of emotion predictions



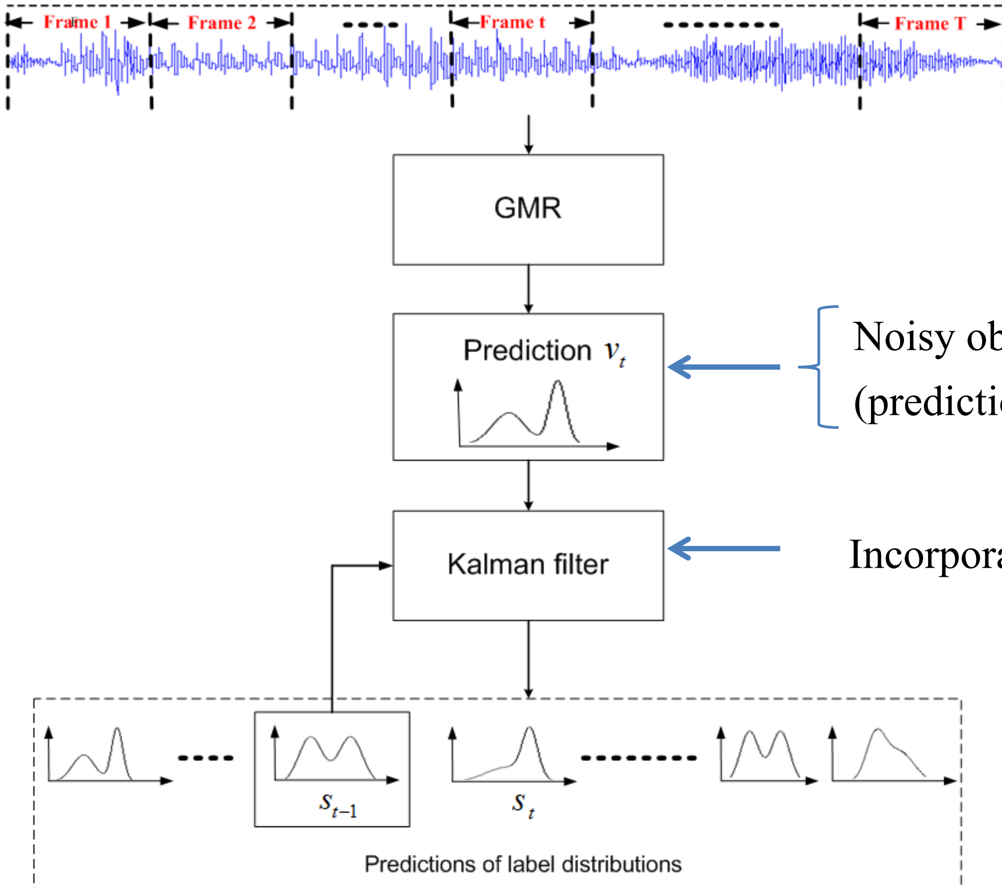

- Standard deviation of six raters correlates with the predicted uncertainty of emotion

## ➢ Limitations

- The assumption of Gaussianity over label distribution may not hold true
- GMR does not model temporal dependencies between frames

# Dynamic multi-rater GMR

- Adopting predicted GMM distribution directly
- Kalman filter is adopted to explore the temporal dependencies

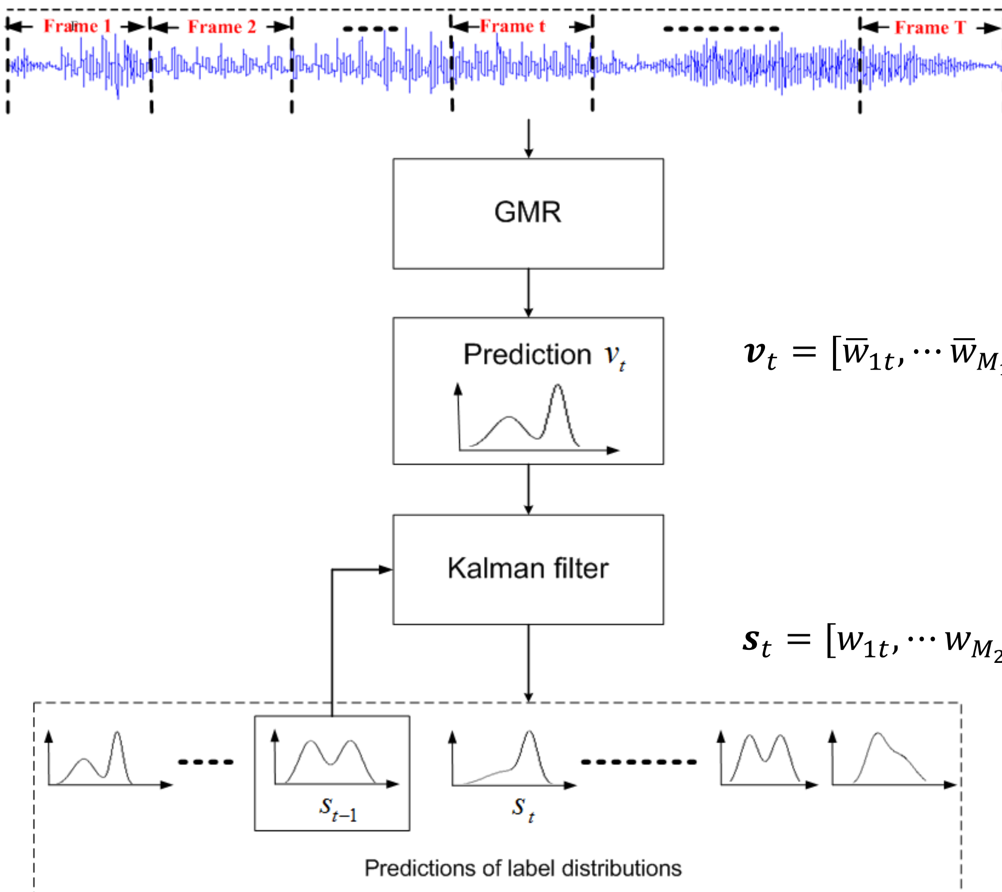



Noisy observations of underlying label distribution (predictions independent of other frames)

Incorporating temporal dependencies

# Dynamic multi-rater GMR

- Adopting predicted GMM distribution directly
- Kalman filter is adopted to explore the temporal dependencies
- Vector representation of GMM distributions is adopted by Kalman filter




$$\boldsymbol{v}_t = [\bar{w}_{1t}, \cdots \bar{w}_{M_1 t}, \bar{\boldsymbol{u}}_{1t}^T, \cdots \bar{\boldsymbol{u}}_{M_1 t}^T, Vec(\bar{\boldsymbol{\Sigma}}_{1t})^T, \cdots Vec(\bar{\boldsymbol{\Sigma}}_{M_1 t})^T]^T$$

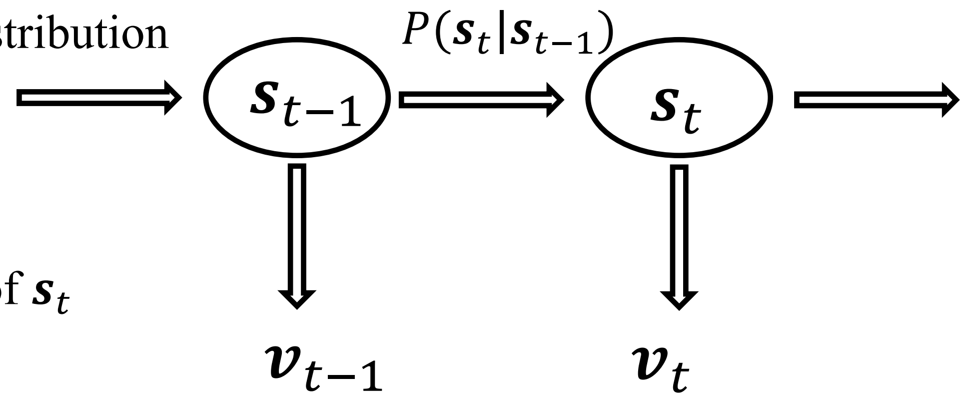
$$\boldsymbol{s}_t = [w_{1t}, \cdots w_{M_2 t}, \boldsymbol{u}_{1t}^T, \cdots \boldsymbol{u}_{M_2 t}^T, Vec(\boldsymbol{\Sigma}_{1t})^T, \cdots Vec(\boldsymbol{\Sigma}_{M_2 t})^T]^T$$

# Dynamic multi-rater GMR

## ➢ Kalman filter

- $\boldsymbol{v}_t$ is treated as the observation of label distribution and $\boldsymbol{s}_t$ is the underlying distribution that depends on the long-term dynamics

$\boldsymbol{s}_t$: underlying label distribution




$\boldsymbol{v}_t$: noisy observation of $\boldsymbol{s}_t$

$$\boldsymbol{s}_t = \boldsymbol{F}\boldsymbol{s}_{t-1} + \boldsymbol{w}_{t-1} \quad (noise\ \boldsymbol{w}_{t-1} \sim N(0, \boldsymbol{Q}))$$

$$\boldsymbol{v}_t = \boldsymbol{H}\boldsymbol{s}_t + \boldsymbol{r}_t \qquad (noise\ \boldsymbol{r}_t \sim N(0, \boldsymbol{R}))$$

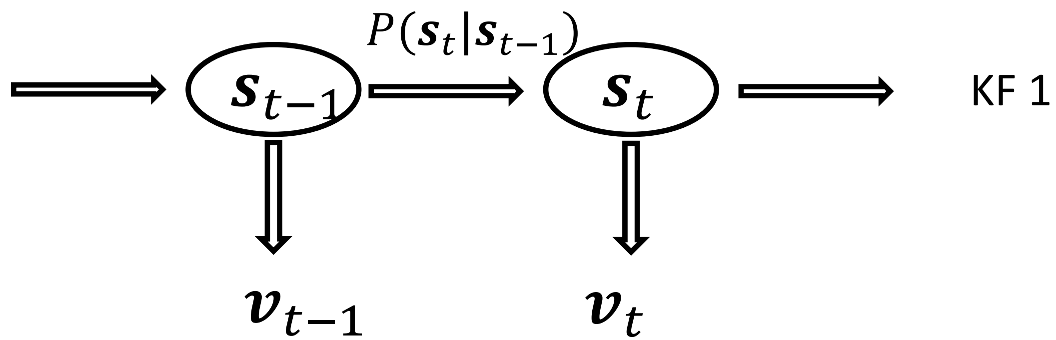
- During training phase, parameters of Kalman filters ($\boldsymbol{F}$, $\boldsymbol{Q}$ ,$\boldsymbol{H}$ and $\boldsymbol{R}$) are estimated where the observations $\boldsymbol{v}_t$ and the ground truth $\boldsymbol{s}_t$ are known.

- During test phase, Kalman filters are utilised to predict the label distribution $\hat{\boldsymbol{s}}_t$ based on the GMR prediction $\boldsymbol{v}_t$ and the prediction of previous frames $\hat{\boldsymbol{s}}_{t-1}$

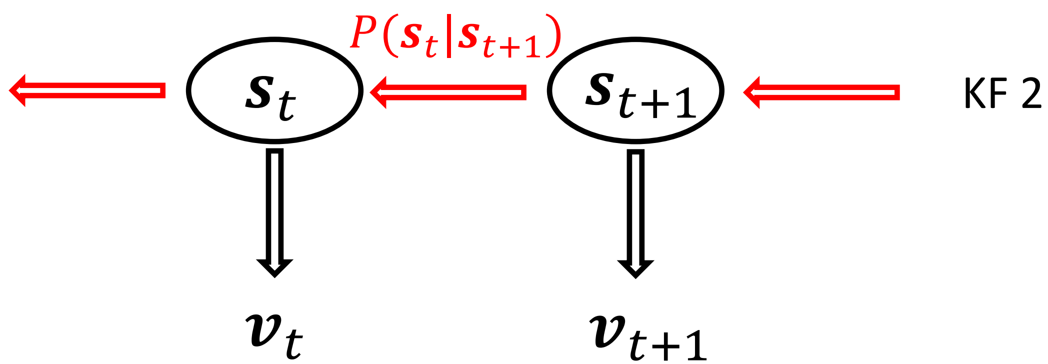
# Dynamic multi-rater GMR

## ➢ Forward and backward Kalman filter

- Forward




- Backward



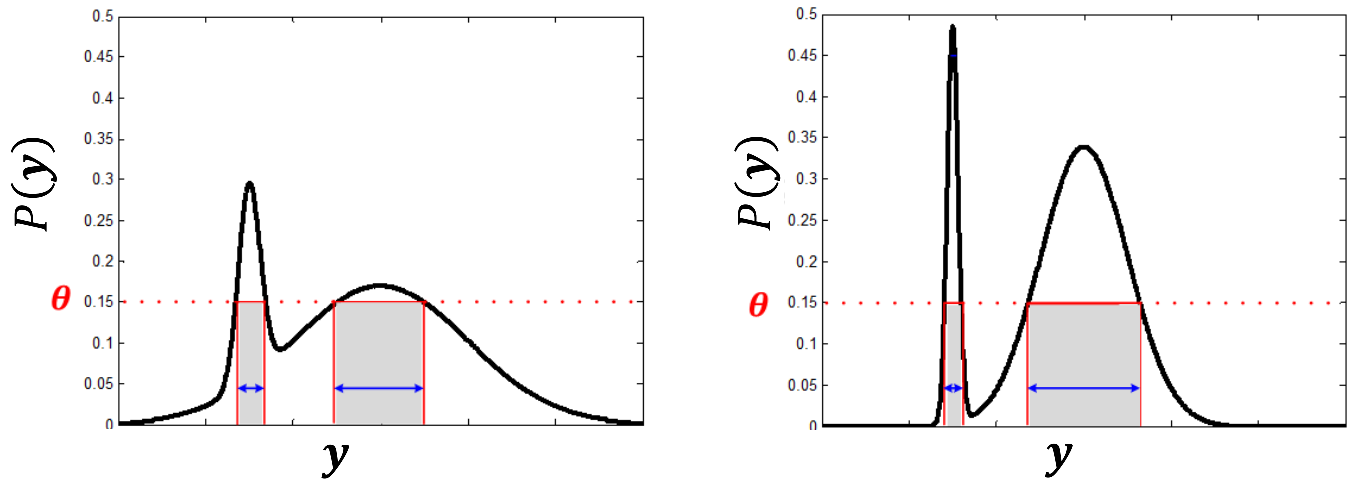

- Final label prediction $\hat{\boldsymbol{s}}_t = \alpha \boldsymbol{s}_t^{KF1} + (1-\alpha)\boldsymbol{s}_t^{KF2}$

$$\hat{\boldsymbol{s}}_t \longrightarrow [w, u, \Sigma]$$

# Measures of Uncertainty

## ➢ Probabilistic uncertainty volume $PUV$

- Probabilistic uncertainty volume estimates the local variability of a distribution




$$PUV_1 \quad > \quad PUV_2$$

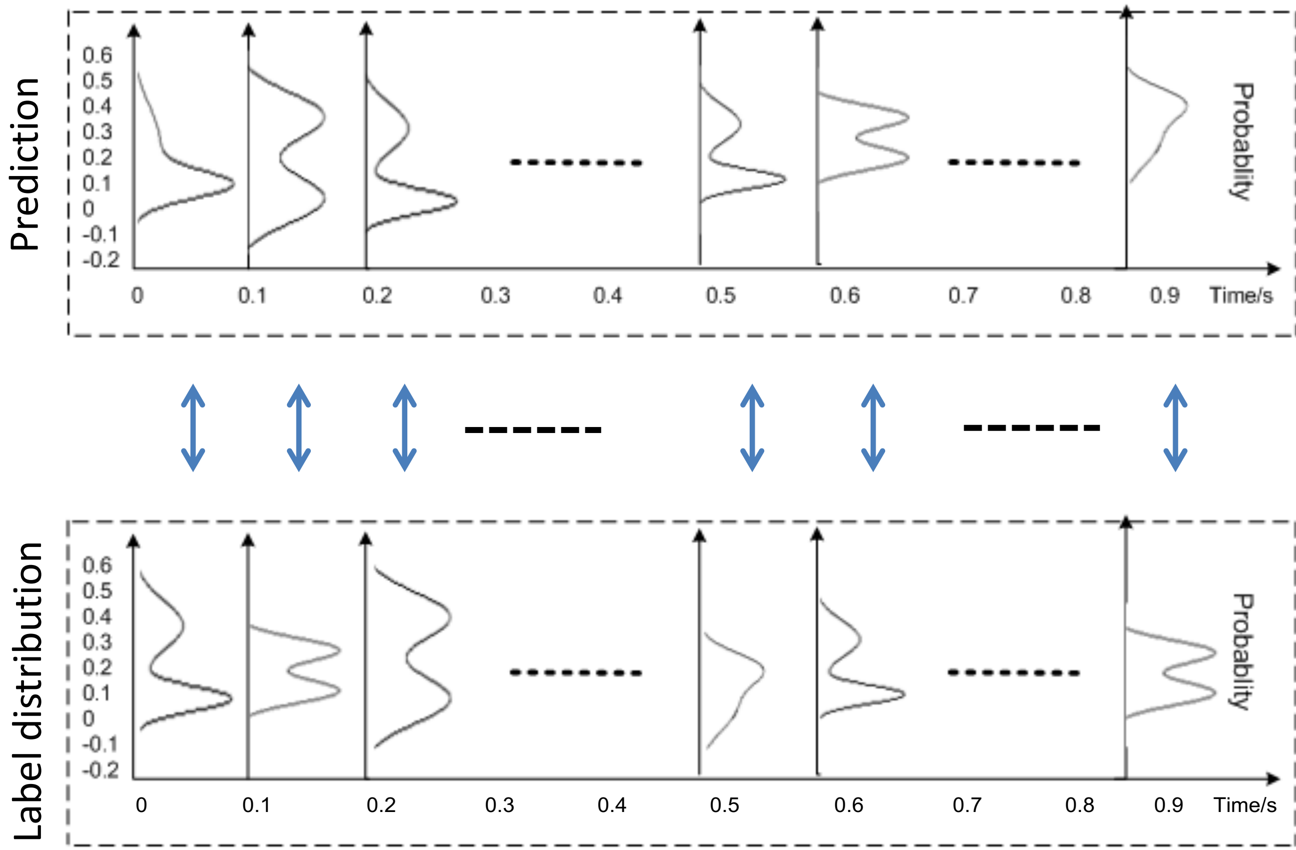
$$PUV_t = \int f(\boldsymbol{y})d\boldsymbol{y}, \qquad f(\boldsymbol{y}) = \begin{cases} 1, & P(\boldsymbol{y}_t) > \theta \\ 0, & P(\boldsymbol{y}_t) \leq \theta \end{cases}$$

- Given threshold $\theta$, $PUV_1$ for a broad GMM (high uncertainty in left side) is larger than $PUV_2$ for a narrow GMM (low uncertainty in right side)

# System Evaluation

- System evaluation focuses on the comparison between predicted and underlying label distributions

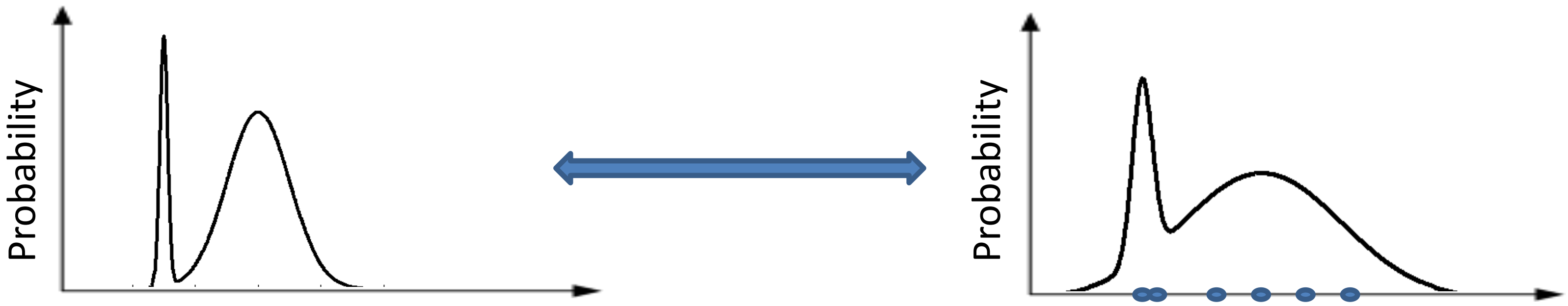
**Predicted by system**




**Inferred from annotations (multiple raters)**

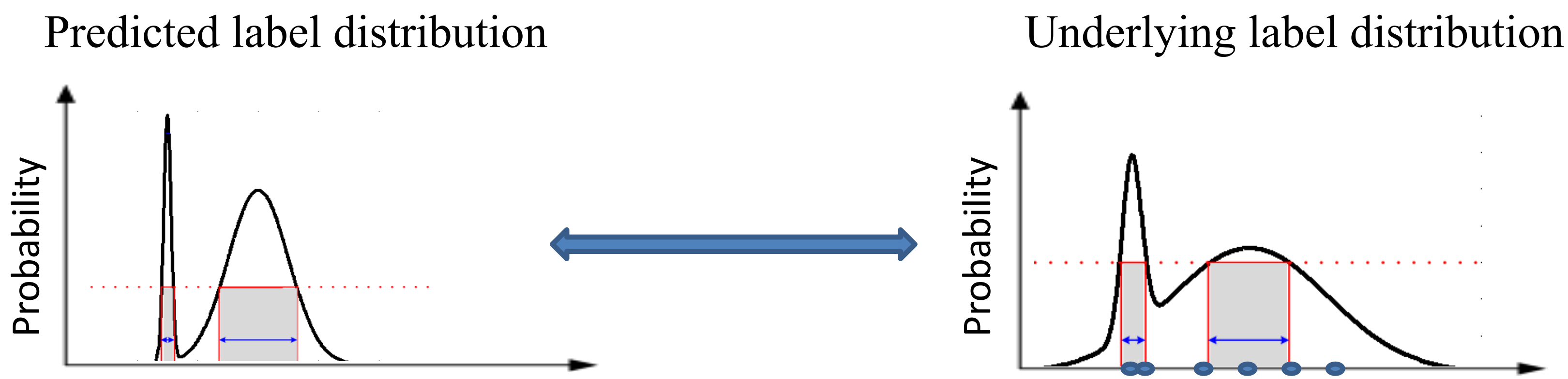
# Evaluation Metrics

Predicted label distribution	Underlying label distribution




- Underlying label distribution (GMM) is time-dependent and estimated in the label space by  6 annotations

# Evaluation Metrics




- Underlying label distribution is time-dependent and estimated in the label space by 6 annotations

- Probabilistic uncertainty volume $PUV$ is estimated for the predicted and underlying label distribution respectively for each frame

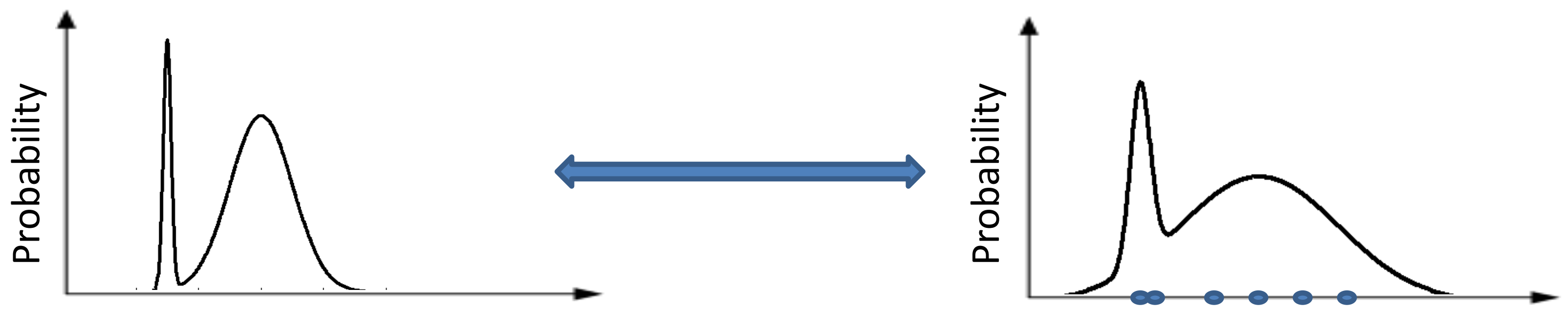
## ➢ Correlation coefficient (CC)

- Pearson's correlation coefficient between probabilistic uncertainty volume estimated from the predicted and the underlying label distribution

- A higher CC indicates better predicted label distributions

# Evaluation Metrics

Predicted label distribution　　　　Underlying label distribution



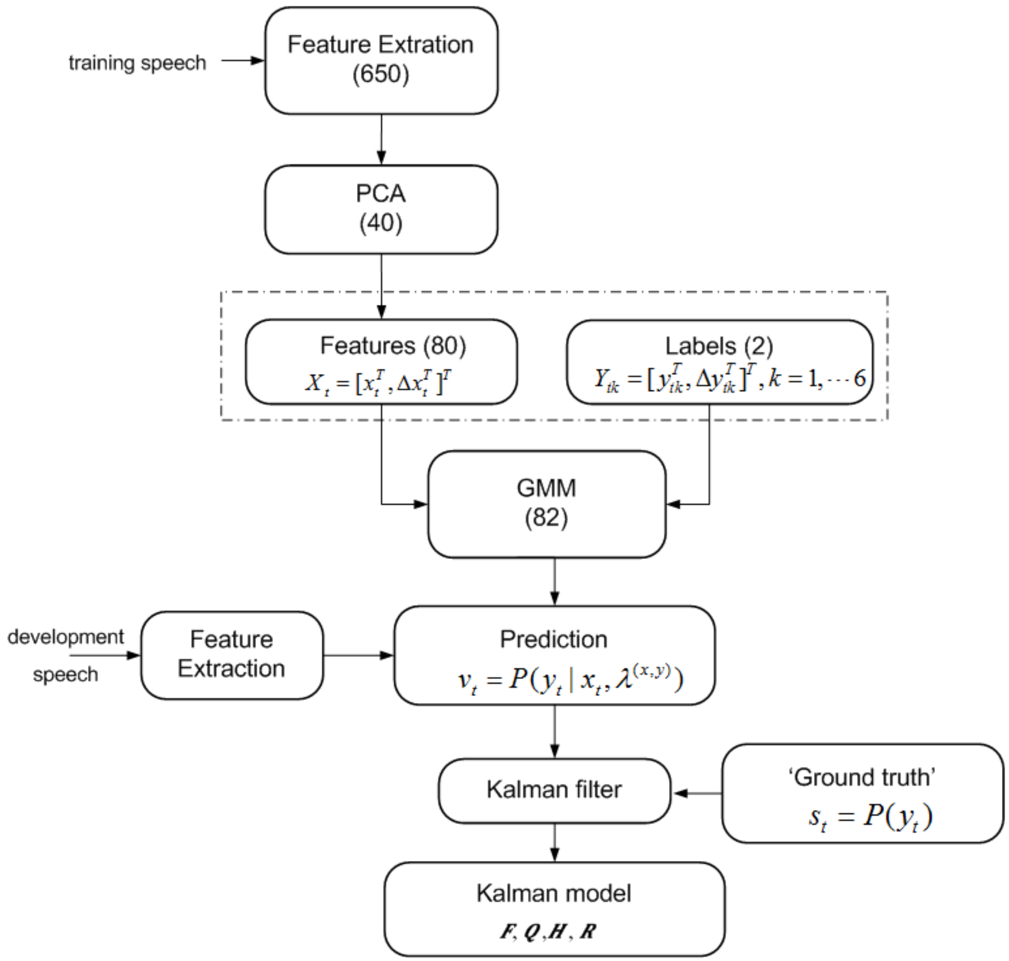

- Underlying label distribution is time-dependent and estimated in the label space by 6 annotations

## ➢ KL divergence

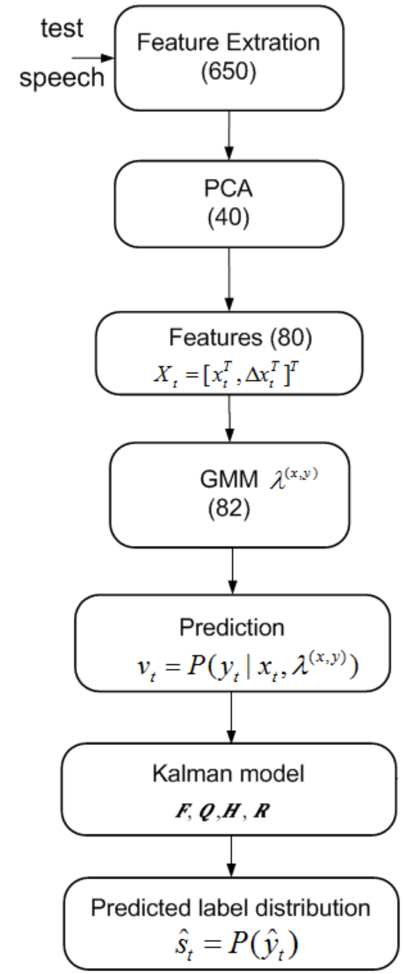
- KL divergence estimates the similarity between the predicted and the underlying label distributions
- A smaller KL divergence indicates better predicted label distributions
- *Median and 25$^{th}$ and 75$^{th}$ percentiles* of KL divergence over entire test dataset are estimated (boxplot)

# Experimental Settings

## ➢ Training phase




## ➢ Test phase




* Github: https://github.com/TingDang90/Dynamic-multi-rater-GMR

# Experimental Settings

## ➢ Experimental settings

- Database: RECOLA (6 annotations)
- Features: 5 functionals applied to 130 LLDs
- PCA : 40 dimensions
- Delays: 2s for arousal and 4s for valence
- GMM mixture number: [2,4,8]
- Linear coefficient of Kalman filter: [0, 1] with a step increase of 0.1

- **Baseline:**

  **---** Multi-rater GMR system

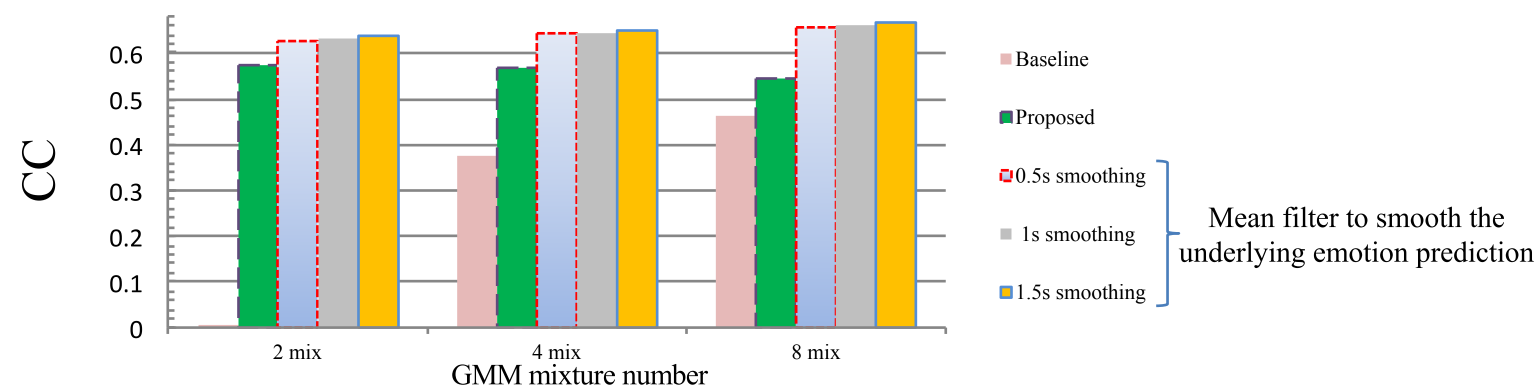
  i. CC between the PUV of predicted Gaussian and PUV of underlying label distribution

  ii. KL between the predicted Gaussian and the underlying label distribution(GMM)

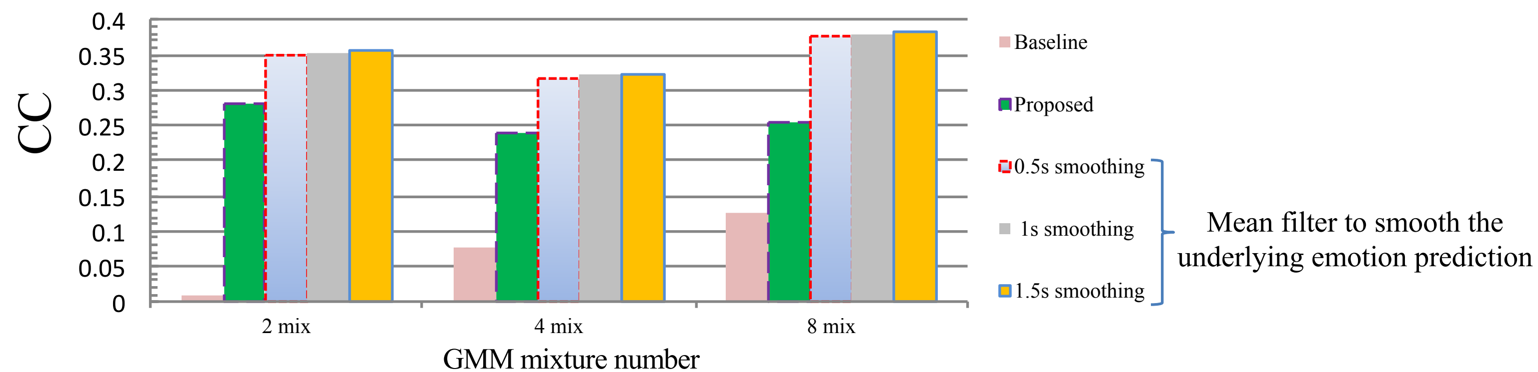
# Experimental Results

## ➢ CC between predicted and true $PUV$

- CC between the $PUV$ of the predicted and underlying label distributions (GMM)




(a) arousal



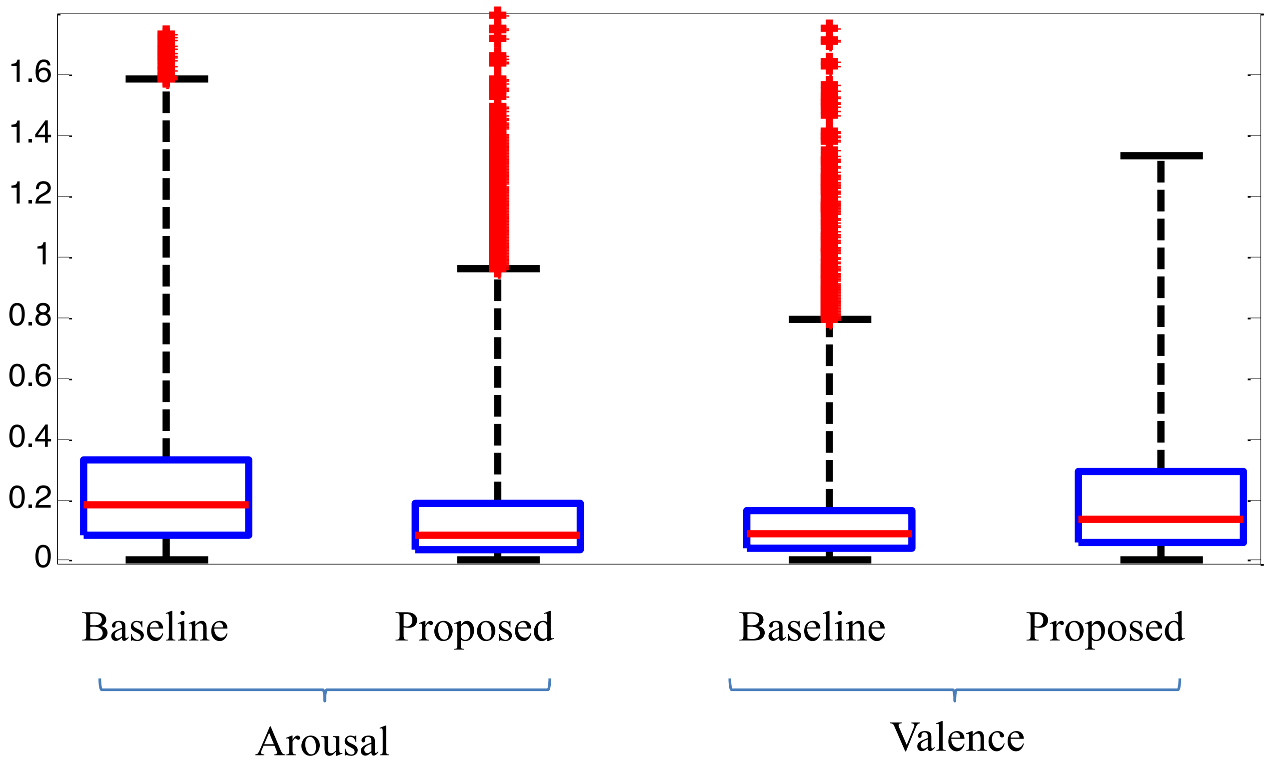

(b) valence

- Incorporating temporal dependencies benefits uncertainty prediction, especially for valence

# Experimental Results

## ➢ KL divergence between predicted and underlying label distributions

- KL between the predicted and underlying label distribution (GMM) is computed




- The proposed system leads to more reliable and smoothed distribution prediction

# Conclusion

- A dynamic multi-rater GMR to predict emotion uncertainty by considering the temporal dependencies is proposed, which is achieved by applying Kalman filters.

- Probabilistic uncertainty volume is introduced as a measure to quantify uncertainty of emotion predictions (GMM).

- The statistics of KL divergence between predicted and underlying label distributions indicate that incorporating temporal dependencies between frames leads to a smoother change in the label distributions

- Future work will focus on relaxing linearity assumption about the evolution of emotion label distributions

# Reference


[1] E. Mower *et al.*, "Interpreting ambiguous emotional expressions," in *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, 2009, pp. 1-8: IEEE.

[2] F. Ringeval *et al.*, "Prediction of asynchronous dimensional emotion ratings from audiovisual and physiological data," *Pattern Recognition Letters,* vol. 66, pp. 22-30, 2015.

[3] R. Lotfian and C. Busso, "Retrieving Categorical Emotions Using a Probabilistic Framework to Define Preference Learning Samples," in *INTERSPEECH*, 2016, pp. 490-494.

[4] F. Eyben, M. Wöllmer, and B. Schuller, "A multitask approach to continuous five-dimensional affect sensing in natural speech," *ACM Transactions on Interactive Intelligent Systems (TiiS),* vol. 2, no. 1, p. 6, 2012.

[5] J. Han, Z. Zhang, M. Schmitt, M. Pantic, and B. Schuller, "From Hard to Soft: Towards more Human-like Emotion Recognition by Modelling the Perception Uncertainty," presented at the ACM MM 2017, Mountain View, 2017.

[6] M. S. Grewal, "Kalman filtering," in *International Encyclopedia of Statistical Science*: Springer, 2011, pp. 705-708.

[7] K. Somandepalli, R. Gupta, M. Nasir, B. M. Booth, S. Lee, and S. S. Narayanan, "Online Affect Tracking with Multimodal Kalman Filters," in *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*, 2016, pp. 59-66: ACM.

[8] T. Toda, A. W. Black, and K. Tokuda, "Voice conversion based on maximum-likelihood estimation of spectral parameter trajectory," IEEE Transactions on Audio, Speech, and Language Processing, vol. 15, no. 8, pp. 2222-2235, 2007.

[9] Z. Huang and J. Epps, "An Investigation of Emotion Dynamics and Kalman Filtering for Speech-based Emotion Prediction," Proc. Interspeech 2017, pp. 3301-3305, 2017.

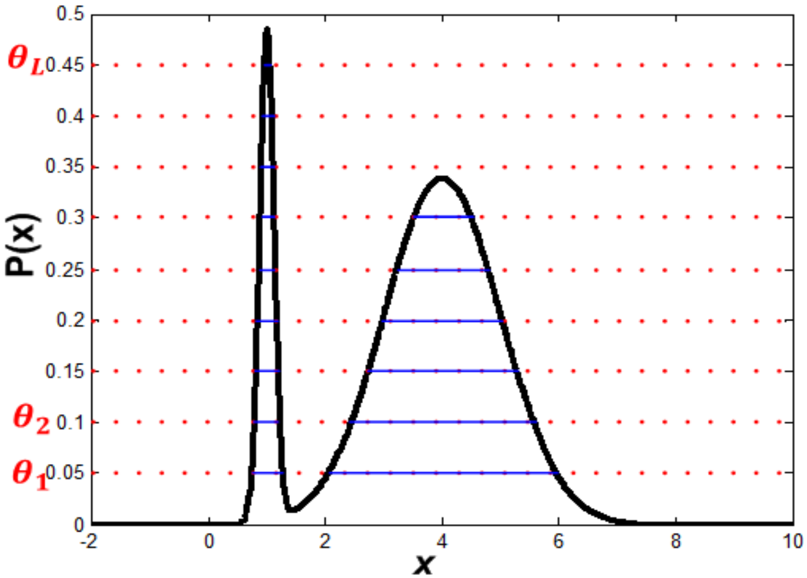
[10] N. Cummins, V. Sethu, J. Epps, and J. Krajewski, "Probabilistic acoustic volume analysis for speech affected by depression," in INTERSPEECH, 2014, pp. 1238-1242.

[11] V. Sethu, J. Epps, and E. Ambikairajah, "Speaker variability in speech based emotion models-Analysis and normalisation," in Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, 2013, pp. 7522-7526: IEEE.

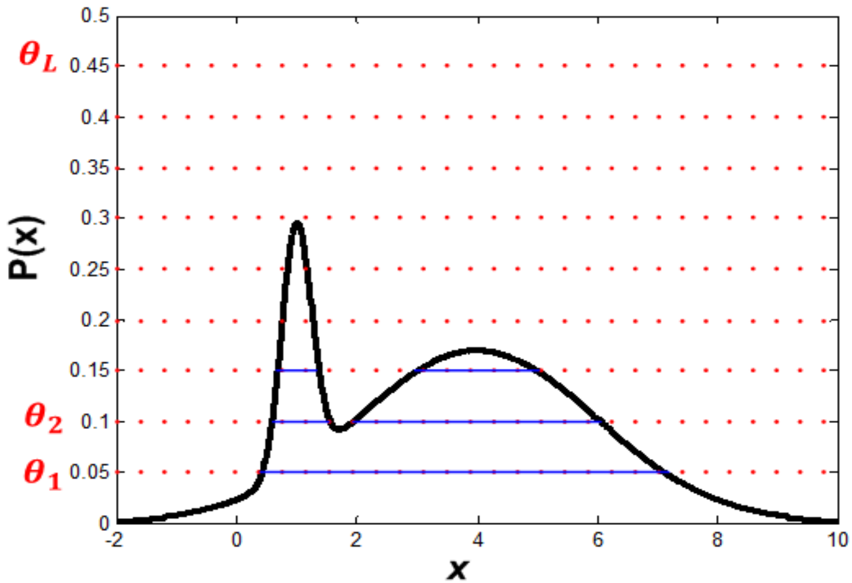
# Thank you

# Thresholds of Probabilistic Uncertainty Volume

Predicted label distribution

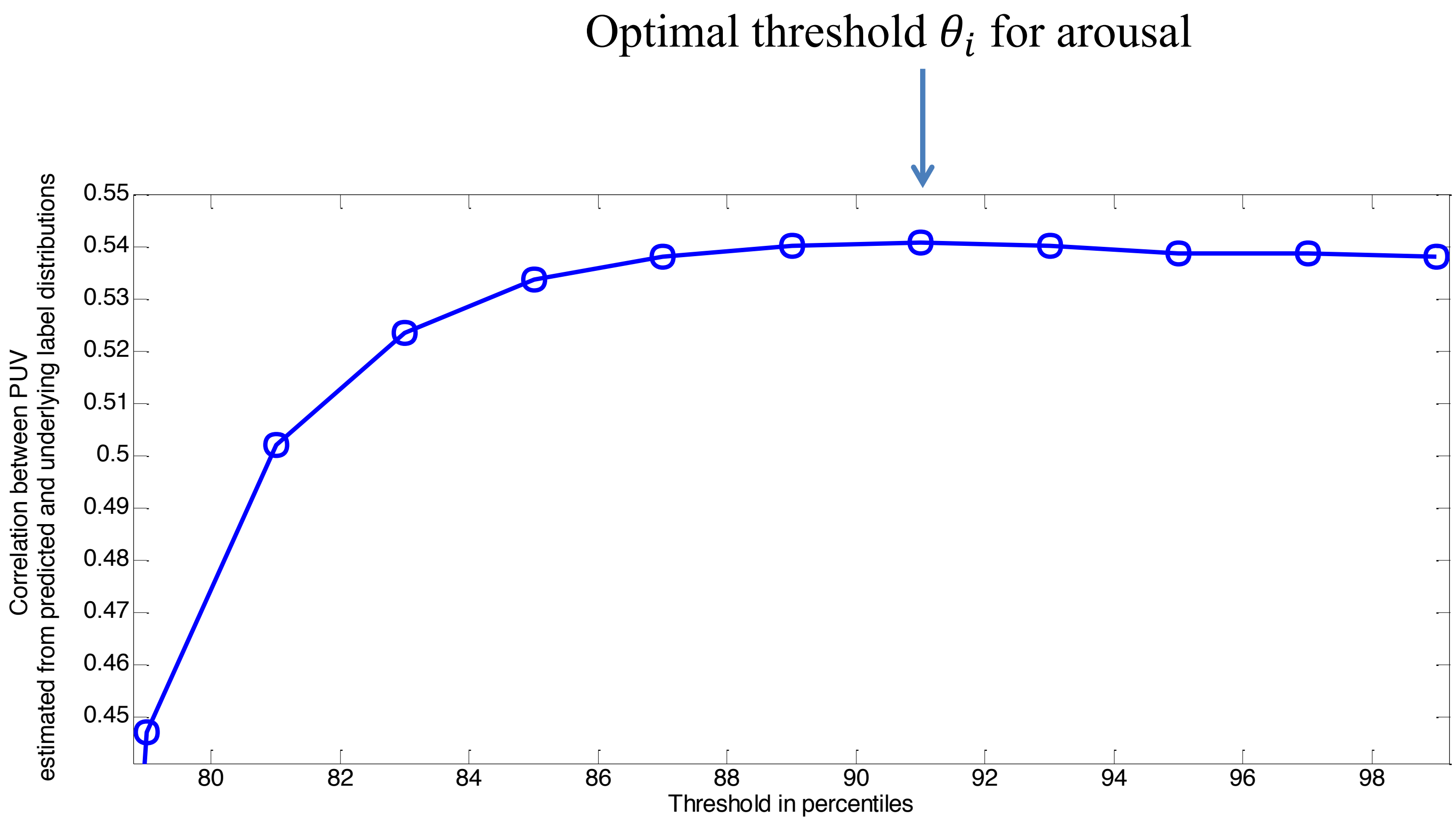



Underlying label distribution

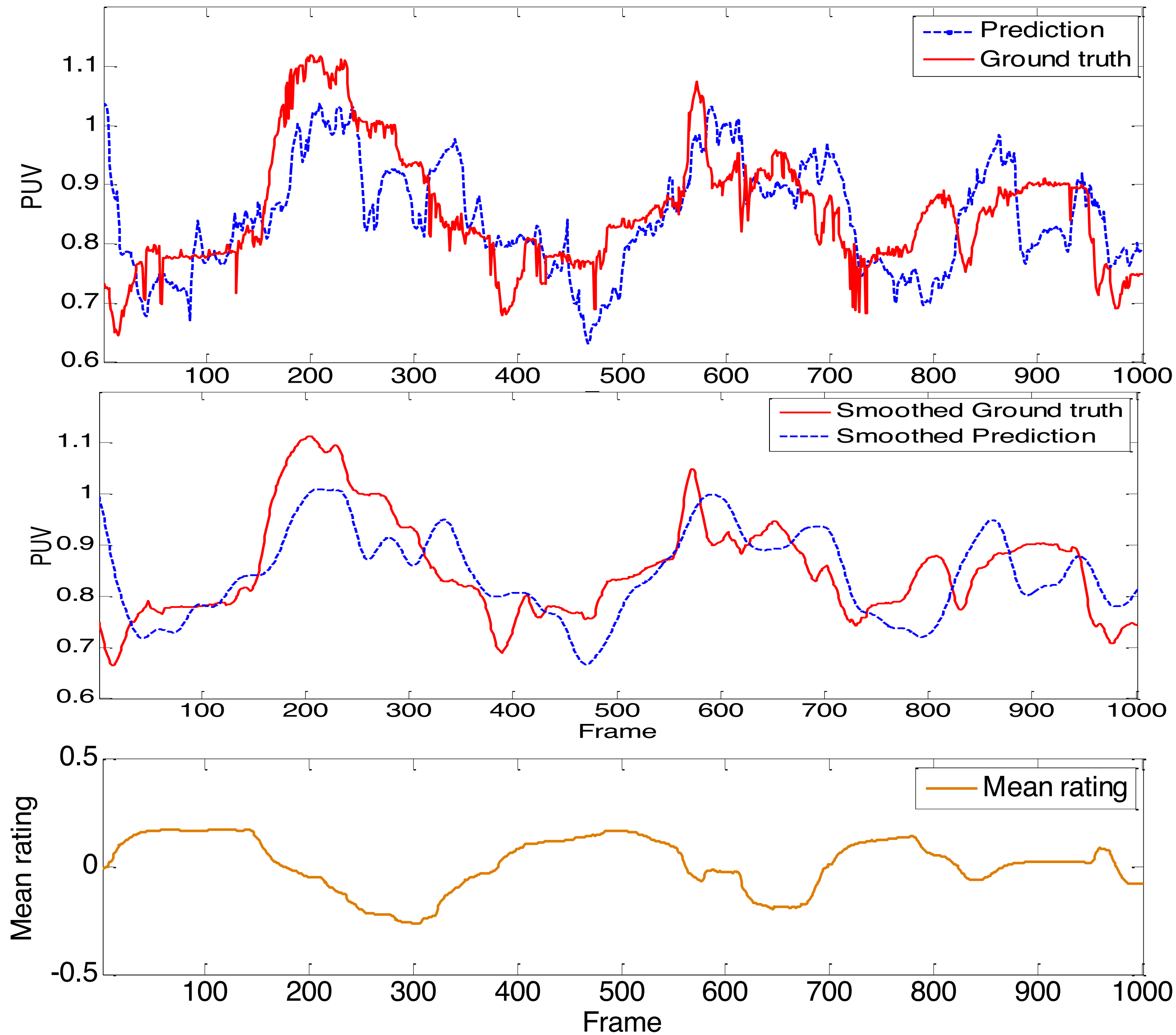



- Thresholds $\theta_i$ are defined in terms of percentiles of all the probabilities calculated by fitting the test features to the GMM models
- The optimal threshold $\theta_i$ is determined experimentally based on the system performance

# CC between PUV from predicted and underlying distributions



# Smoothness of PUV from underlying label distribution



# KL divergence

- Symmetric KL divergence is utilised, with a larger KL divergence indicating a greater separation between them.

- Specifically, a Monte-Carlo estimate of the symmetric KL divergence proposed in [11] is utilised to quantify the separation between two distributions.

$$I_{SKL}(P_1, P_2) = \frac{1}{2}\left|\int_x P_1(\boldsymbol{x}) In \frac{P_1(\boldsymbol{x})}{P_2(\boldsymbol{x})} dx + \int_{\boldsymbol{x}} P_2(\boldsymbol{x}) In \frac{P_2(\boldsymbol{x})}{P_1(\boldsymbol{x})} d\boldsymbol{x}\right| \qquad (3.1)$$

# Experimental Results

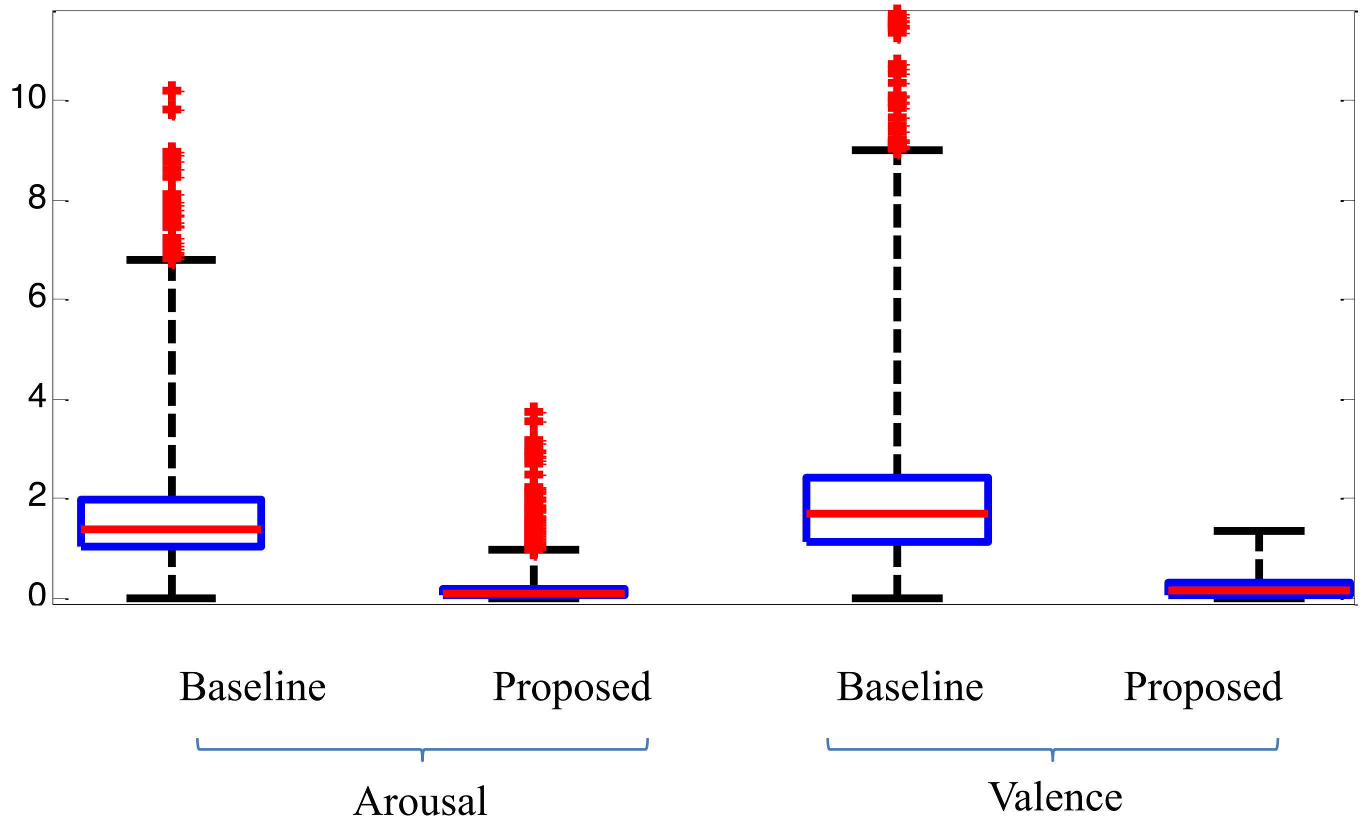
- **KL divergence between predicted and underlying label distributions**

| | Arousal | | Valence | |
|---|---|---|---|---|
| | Proposed | Baseline | Proposed | Baseline |
| **Mean** | 0.1439 | 1.6872 | 0.2085 | 1.8628 |
| **SD** | 0.1818 | 7.2714 | 0.2044 | 1.1236 |

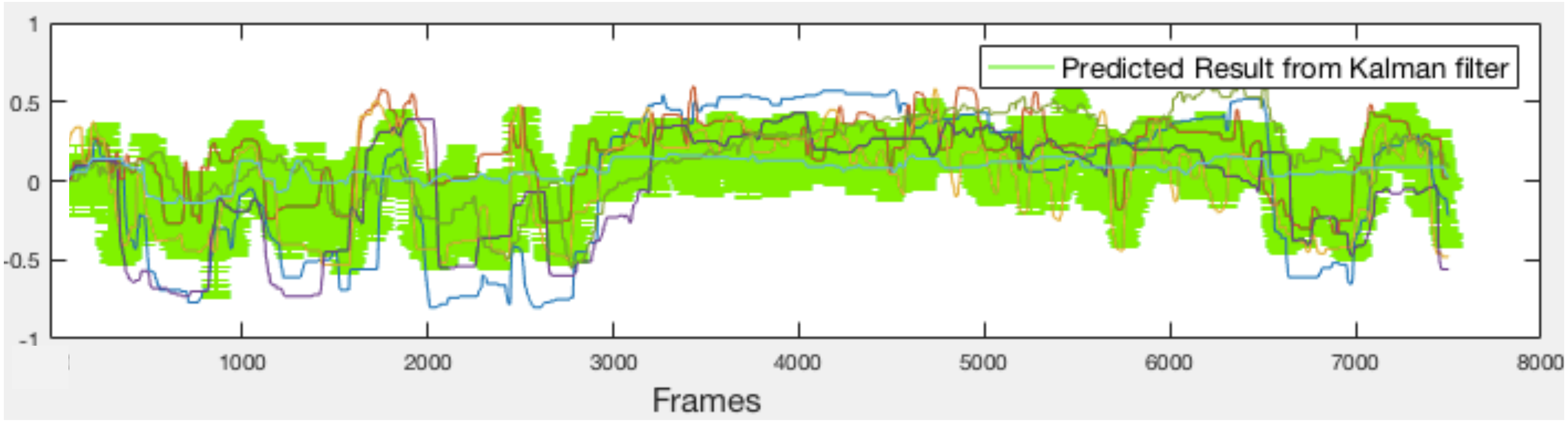
--- Baseline means the KL diverenge calculated between predicted and underlying GMM distributions.

--- The proposed system leads to more reliable and smoothed distribution prediction

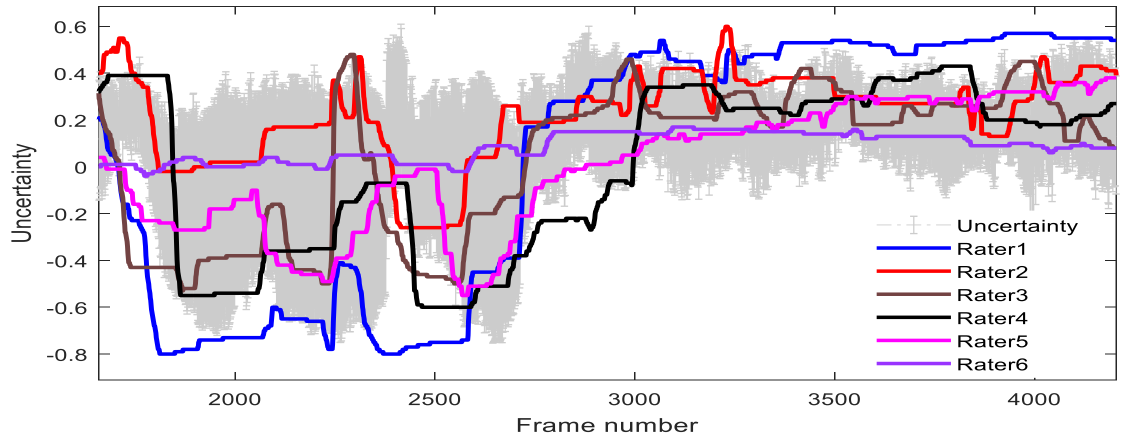
# KL divergence



# Uncertainty Prediction using Kalman filters



# Uncertainty Prediction using Kalman filters



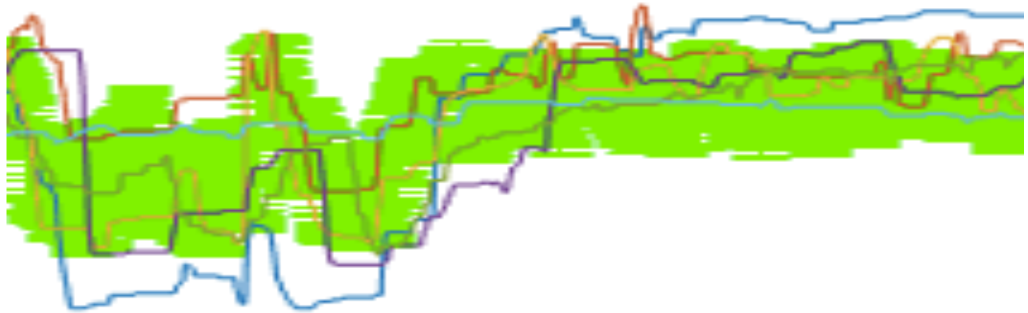
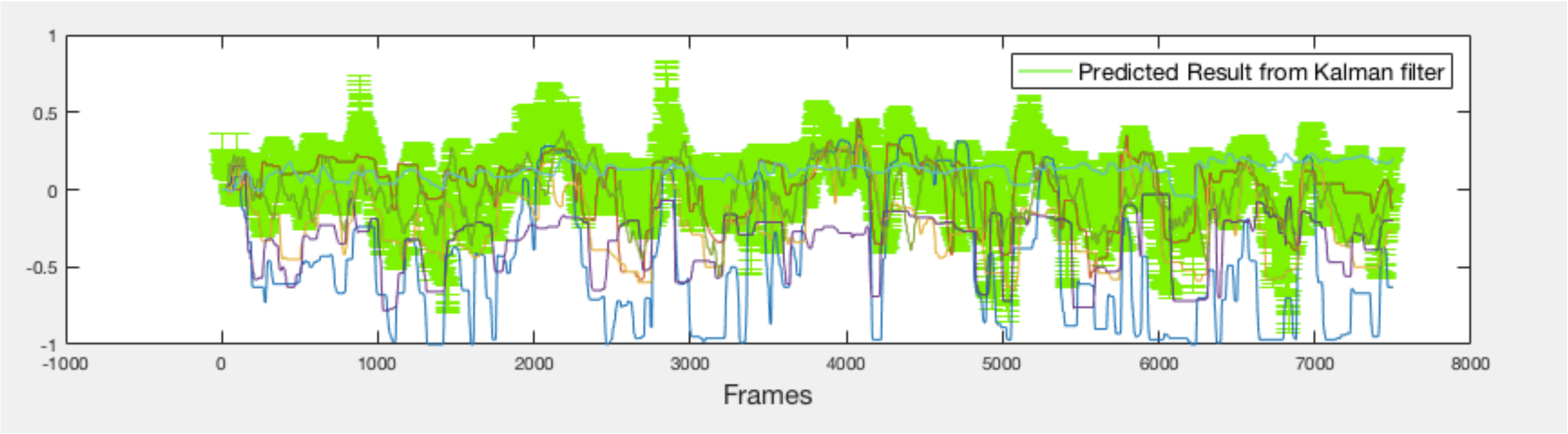



Figure. 25-75% quartile plotted as error bar, with 6 true annotations overlaid.
utterance 2 in dev set
Yellow: predicted GMM(ESN) ; Cyan: assumed 'ground truth'; Green: predicted GMM(Kalman filter)

Left: utterance 4 in dev set;



# Kalman filter

$$P(\boldsymbol{s}_t|\boldsymbol{s}_{t-1}) = N(\boldsymbol{s}_t; \boldsymbol{F}\boldsymbol{s}_{t-1}, \boldsymbol{Q})$$

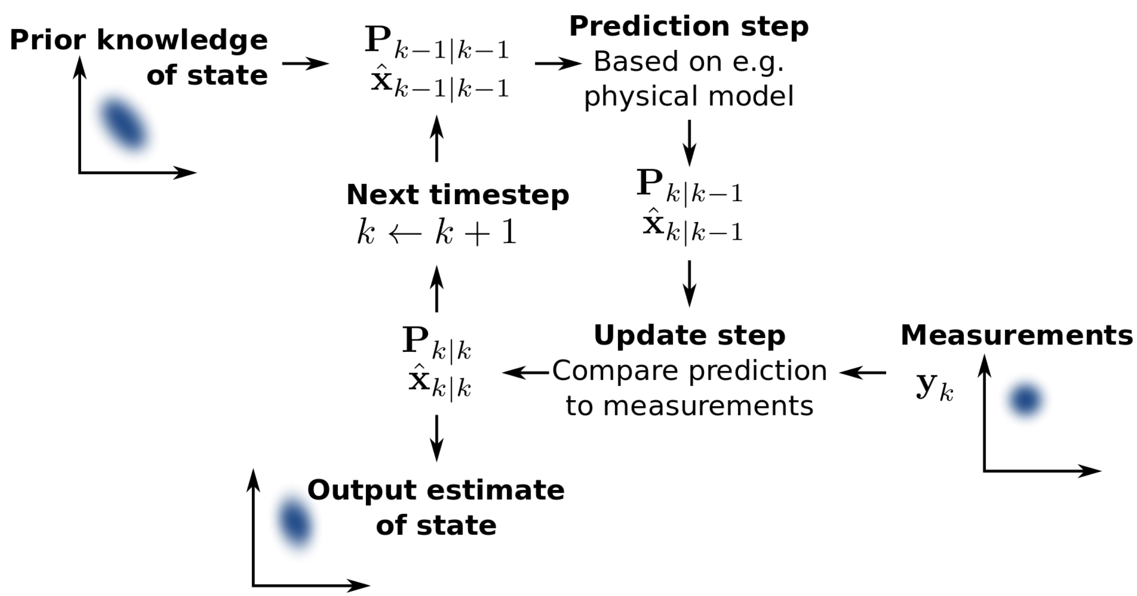
$$P(\boldsymbol{v}_t|\boldsymbol{s}_t) = N(\boldsymbol{v}_t; \boldsymbol{H}\boldsymbol{s}_{t-1}, \boldsymbol{R})$$

$$\boldsymbol{F} = (\boldsymbol{A}^T\boldsymbol{A} + \lambda\boldsymbol{I})^{-1}\boldsymbol{A}^T\boldsymbol{B}$$

$$\boldsymbol{Q} = cov(\boldsymbol{B} - \boldsymbol{A}\boldsymbol{F})$$

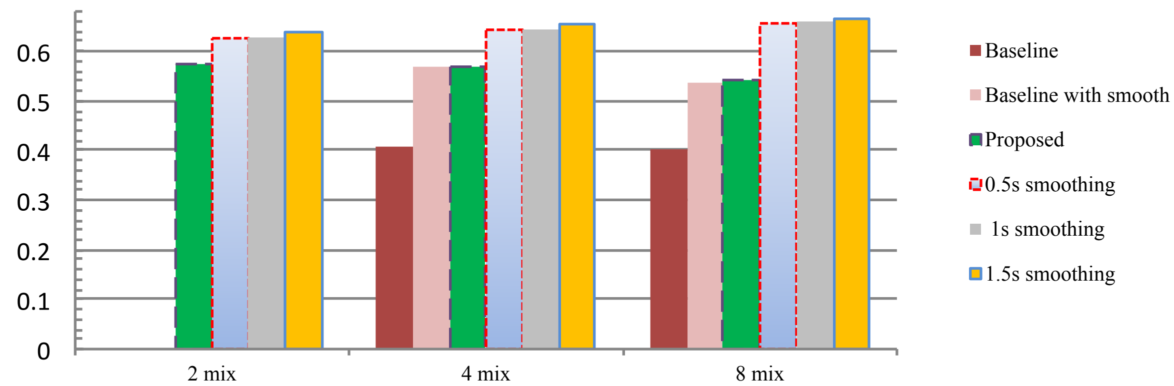
$$\boldsymbol{H} = (\boldsymbol{C}^T\boldsymbol{C} + \lambda\boldsymbol{I})^{-1}\boldsymbol{C}^T\boldsymbol{D}$$

$$\boldsymbol{R} = cov(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{H})$$

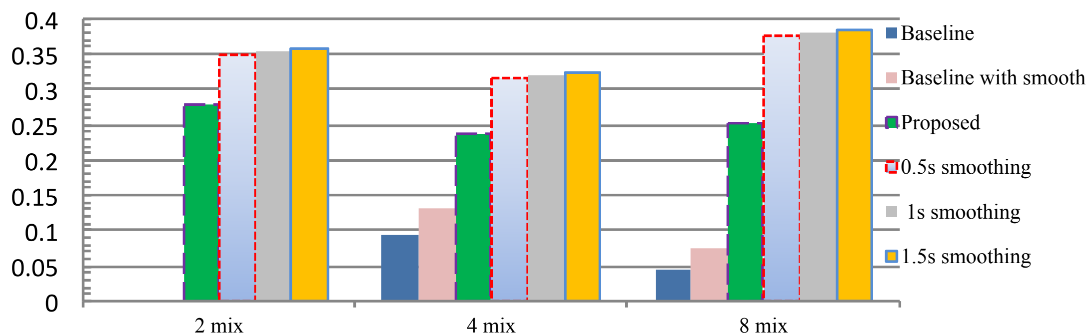


CC between the standard deviation (SD) of predicted Gaussians and PUV (ground truth)

| | Arousal | Valence |
|---|---|---|
| 2 mix | 0.0050 | 0.008 |
| 4mix | 0.3726 | 0.075 |
| 8mix | 0.4632 | 0.1243 |
| CC | 0.2392 | 0.0512 |




(a) arousal




(b) valence