# Seminar
## Facial Image Processing and Analysis

**Emotion Recognition**
**Lecturer: Sócrates Ponce**
**Karlsruhe, 26. January 2012**

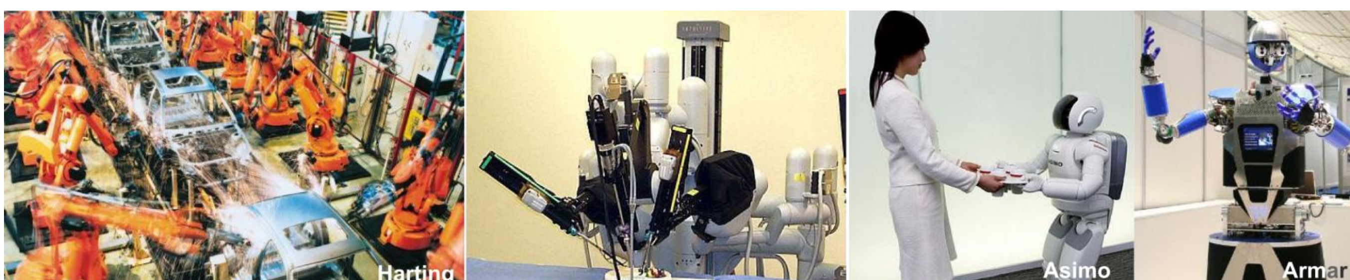
Institute for Anthropomatics, Computer Science Department

# Overview

- Introduction
  - Why is the Recognition of emotions important?
  - What are Emotions?
  - Emotions Theories
  - Ways in which humans express their emotions
- OA - RVM Regression for Dimensional and Continuous Emotion Prediction
  - Related work
  - Proposed idea
  - Experiments
  - Results
- Conclusion

# Introduction

- Influence of computers, robots and related devices has been enormously expanded
- Extension and complexity of computer based solutions increase
- Smooth and easy to used interfaces are needed.



# Why is the Recognition of emotions important?

- Emotions are an important part of human lives.
- Emotions affect and influence the behaviour of humans. For Example:
  - Their learning process
  - Their decision making process
  - Their interaction with other humans beings
- Emotions are researched in various scientific disciplines, e. g.:
  - Neuroscience
  - Psychology
  - Cognitive science
  - Computer science

# Why is the Recognition of emotions important?

(for the field of computer science)

- A way to improve and ease the use of extensive and complex application in a steady growing diversity of environments
  - Try to create solutions that foresees and take into account the [emotional] state of the human operator.

- A way to test the models proposed by psychology, neuroscience, cognitive sciences and computer science as well

# What are Emotions?

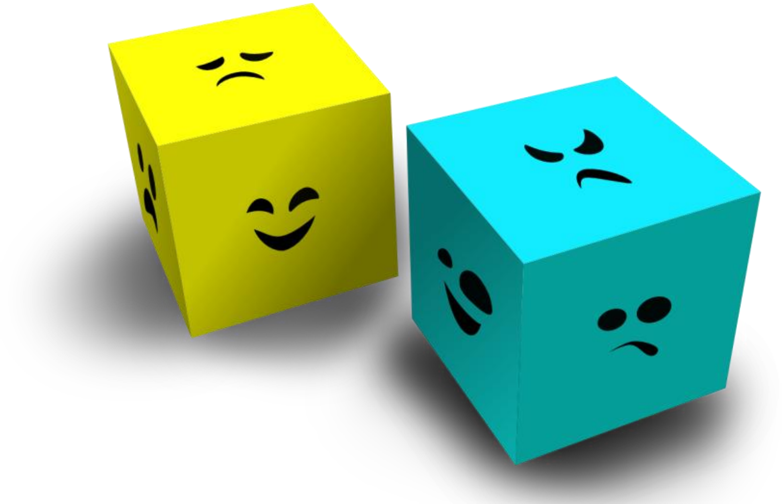
- **Emotion** from the French word **émouvoir**.
  - This based on the Latin *emovere,* where *e*- (variant of *ex*-) means "without" + *movere* means "move".

- **Emotion:** "a complex psychophysiological experience of an individual's state of mind as interacting with biochemical (internal) and environmental (external) influences." [Myers, David G., 2004]


Institute for Anthropomatics, Computer Science Department

# Emotion Theories

- A theoretical explanation about emotions began at least with philosophers like:
  - Plato, Aristotle, the stoics (ancient Greece).
  - Descartes, Spinoza, Hume developed more sophisticated theories.
- With the refinement of the scientific method, new theories raised and those tend to be informed with data obtained from empirical research.



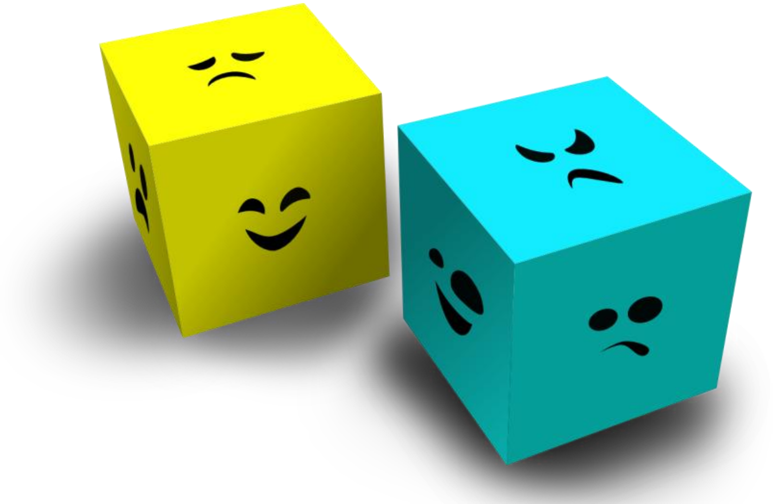

Institute for Anthropomatics, Computer Science Department

# Emotion Theories

- Emotions theories also present **approaches** intended to model the different emotions and link the data generated by humans.

- According to research in psychology three major approaches can be distinguished:

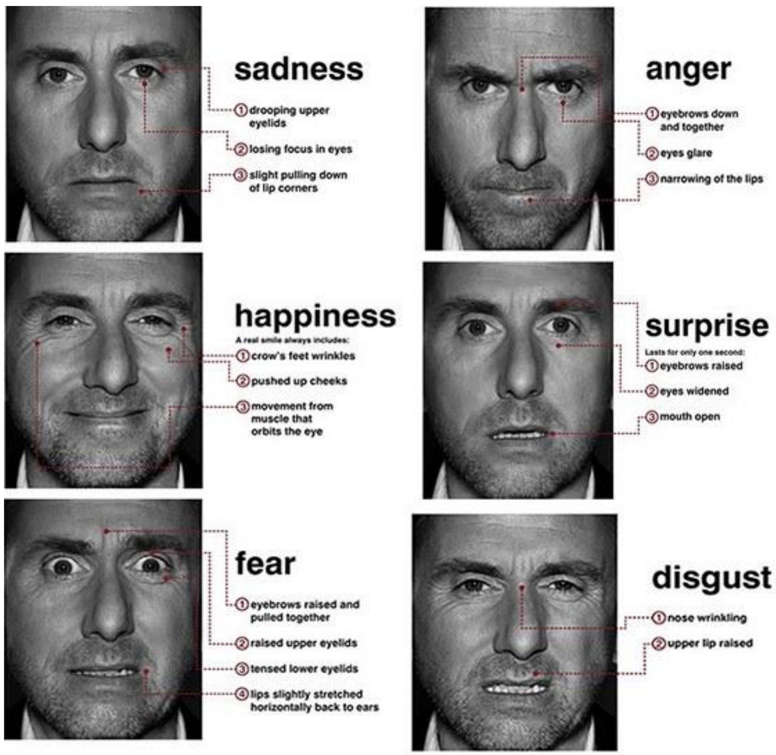
  - Categorical
  - Dimensional
  - Appraisal-based



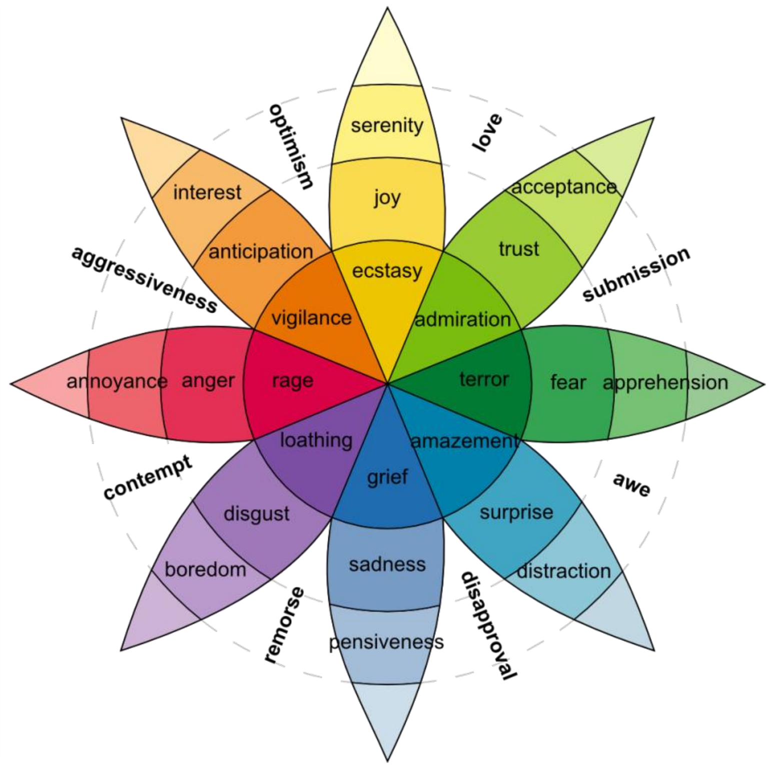

Institute for Anthropomatics, Computer Science Department

# Categorical approach

- Propose the existence of a small number of emotions
- This emotions are basic and hard-wired in our brain
  - → This emotions are universally recognised.[Ekman, P. (1975)]



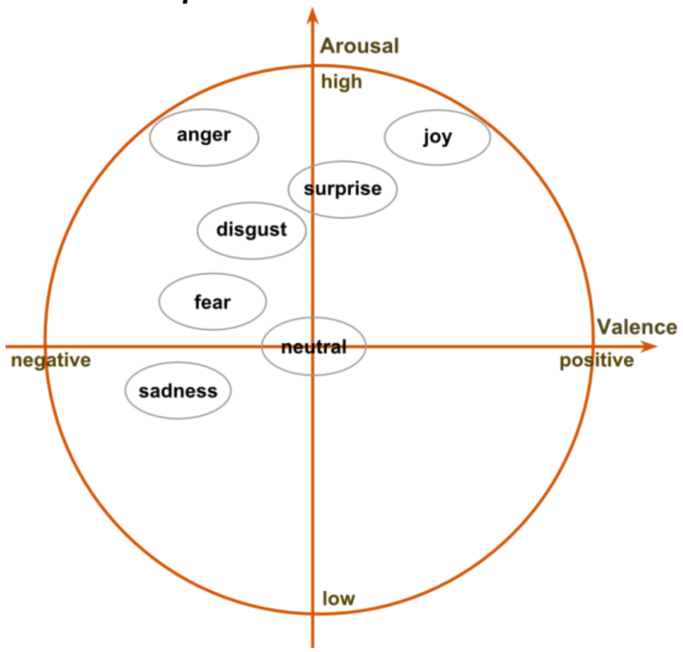

The original 6 basic emotions proposed by P. Ekman

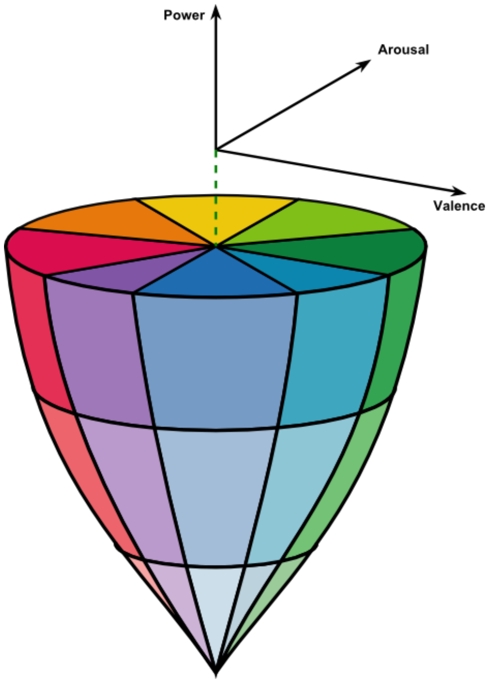


# Dimensional approach

- A number of researchers showed that in every day interactions people exhibit non-basic, subtle and rather complex affective states.
- Affective states are not independent from one another; they are related to one another in a systematic manner
- Most widely used dimensional model:

  *Circumplex of Affect* [Russel, J. A. (1980)]




V-valence, A-arousal, P-power emotion space

# Appraisal-based approach



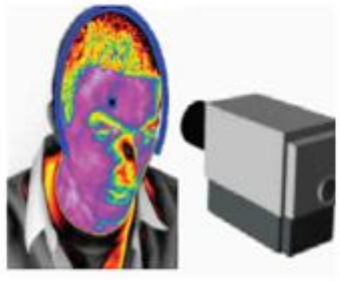
- Emotions are generated through continuous, recursive subjective evaluation of both own internal state and the state of the outside world.

- How to use the appraisal-based approach for automatic measurement of affect is an open research question.

# Ways in which humans express their emotions

- Emotions trigger a flow of signals (cues) in human beings.
- They can divided in:
  - Bio-signals e.g.:
    - Galvanic skin response
    - Electromyography
    - EEG
    - Thermal signature
  - Audio Signals
    - Fundamental frequency
    - Mean intensity
    - Speech rate
  - Visual Signals
    - Configuration of certain features
    - Movement patterns.



# OA - RVM Regression for Dimensional and Continuous Emotion Prediction

- Most dominant techniques used in machines learning and computer vision:
  - Support Vector Machines (SVM)
  - Relevance Vector Machines (RVM)
  - Gaussian Process (GP)

- Many problems expose an inherent dependency amongst the output dimensions.

- An affective state can be described by  a number of latent dimensions.

# Related Work on Dimensional and Continuous Emotion Prediction

Karlsruhe Institute of Technology

- Works focused on predicting continuous and real values are few.

  →The ones in existence use speech features and following mathematical models:
  - Recurrent neural networks (Long Short-Term Memory) and SVR[Wollmer, et al. 2008]
  - SVR, k-NN and a fuzzy logic estimator[Kanluan, Grimm Kroschel, 2008]]

  Other related mathematical models:
  - Kernel Dependency Estimation (KDE) [Weston, 2002]
  - Reformulation of KDE without KPCA [Cortes, et al. 2005]
  - Kernel Ridge Regression (KRR)
  - Twin GP model

- Non of these work explore input-output associations and spatio-temporal dependencies between the output vectors

# Proposed Idea

- RVM regression
  - Our goal is to learn the functional

$$t_i = \mathbf{w}^T \phi(\mathbf{x_i}) + \epsilon_i$$

where:

- $\epsilon_i$ are assumed to be independent Gaussian samples with zero mean and $\sigma^2$ variance.
- $\phi$ is a typically non-linear projection of the input features, $\mathbf{x_i}$

because

- Many problems expose an inherent dependency amongst the output dimensions

  → extends the traditional RVM regression proposing an Output-Associative RVM (OA-RVM) regression.

# Output-Associative RVM regression

- OA-RVM regression

  We introduce:

  $$t_i = \mathbf{w}^T \phi_w(\mathbf{x_i}) + \mathbf{u}^T \phi_u(\mathbf{y}_\mathrm{i}^\mathrm{v}) + \epsilon_i$$

  where:

  - $\mathbf{y}_\mathrm{i}^\mathrm{v}$ is a vector of multi-dimensional output over a temporal window of $[i - v, i + v]$
  - $\mathbf{x_i}$ are called the input features
  - $\mathbf{y}_\mathrm{i}^\mathrm{v}$ are called the output features

- The goal now becomes learning not only the set of weight for the input features, but also the set of weight for the output features along with the noise estimate, $(\epsilon_i)^2$.

# OA-RVM: Inference

- Before any prediction can take place, we to have maximise:

$$P\left(\mathbf{t}|\alpha,\zeta,\sigma^{\mathbf{2}}\right) = \int P\left(\mathbf{t}|\mathbf{w},\mathbf{u},\sigma^{\mathbf{2}}\right) P\left(\mathbf{w},\mathbf{u}|\alpha,\zeta\right) d\mathbf{w}\, d\mathbf{u}$$

  - where:
    - $\alpha, \zeta$ are vectors of hyperparameters, that describe the weight distribution of $\mathbf{w}, \mathbf{u}$.

# Output-Associative RVM regression

- After we *infer* the desired parameters, the prediction step can be carried out:
  - Given a new input data $\mathbf{x}_*, \mathbf{y}_*^{\mathbf{v}}$, we want to calculate $t_*$

$$t_* = \mu_{wu}^T \left[\phi_w(\mathbf{x}_*) | \phi_u(\mathbf{y}_*^{\mathbf{v}})\right]$$

    where:

    - $\mu_{wu}$ contains the weights for the input and output relevance vectors

  - The basis matrix for a new set of test points should now contain the distances from the new test input[output] features to all input[output] feature relevance vectors

# Experiments

- Two types of experiments were carried out:
  - Subject-dependent
  - Subject-independent
- The Sensitive Artificial Listener (SAL) Database were used
  - Segments capturing transitions to an emotional state and back were generated.
  - 61 positive and 73 negative segments were used (aprox. 30 000 video frames)

# Results: Sparsity

- RVM vs. OA-RVM:

TABLE I
SUBJECT-DEPENDENT SPARSITY COMPARISON

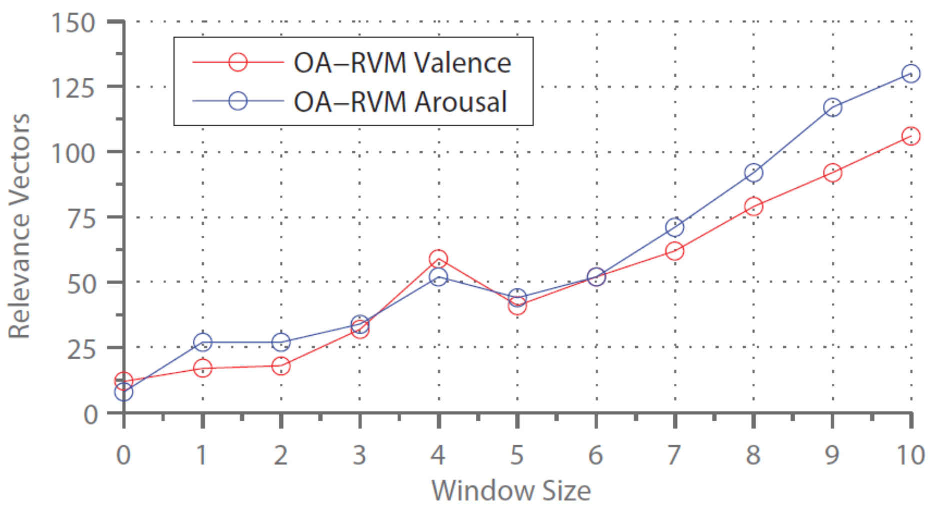
| | Valence$_{RV}$ | | | Arousal$_{RV}$ | | |
|---|---|---|---|---|---|---|
| | RVM | OA-RVM | RMSE | RVM | OA-RVM | RMSE |
| Positive | 267 | 10 | 0.23 | 270 | 12 | 0.22 |
| Negative | 245 | 10 | 0.23 | 244 | 13 | 0.36 |

A smaller set or relevance vectors (RV) implies a less complex model, with a reduced risk of overfitting

TABLE II
SUBJECT-INDEPENDENT SPARSITY COMPARISON

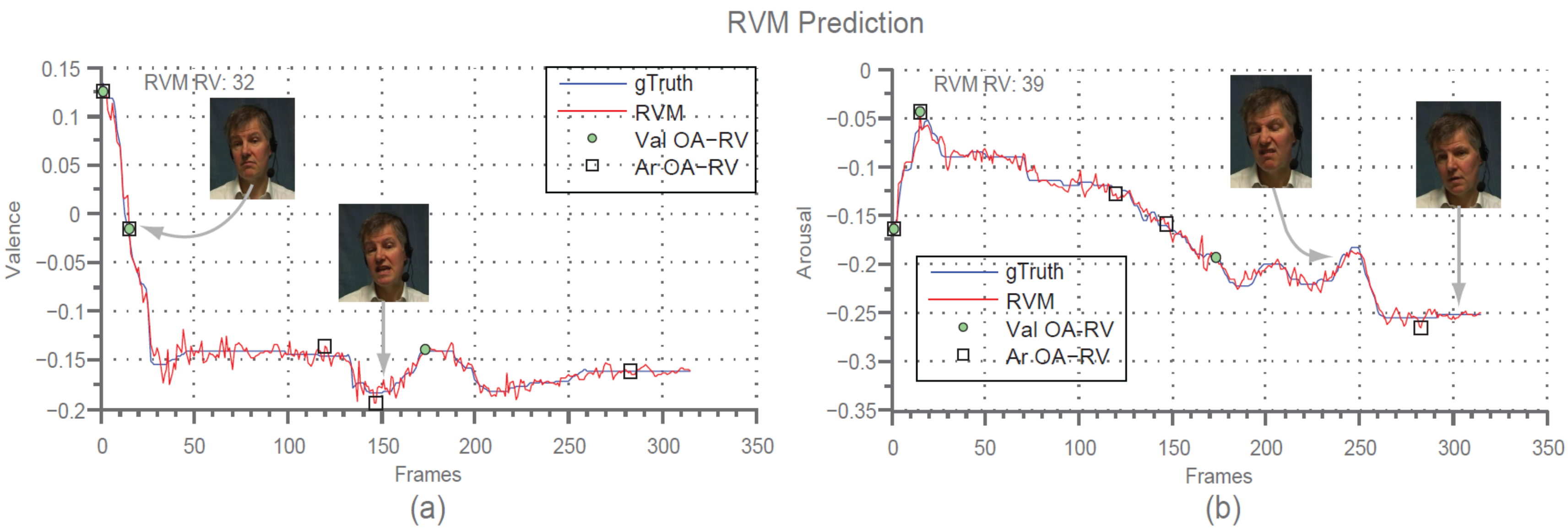
| | Valence$_{RV}$ | | | Arousal$_{RV}$ | | |
|---|---|---|---|---|---|---|
| | RVM | OA-RVM | RMSE | RVM | OA-RVM | RMSE |
| Positive | 485 | 10 | 0.2 | 495 | 11 | 0.15 |
| Negative | 394 | 21 | 0.19 | 417 | 29 | 0.36 |

A larger window complicates the model an increase the number of RVs needed.



# Results: Prediction

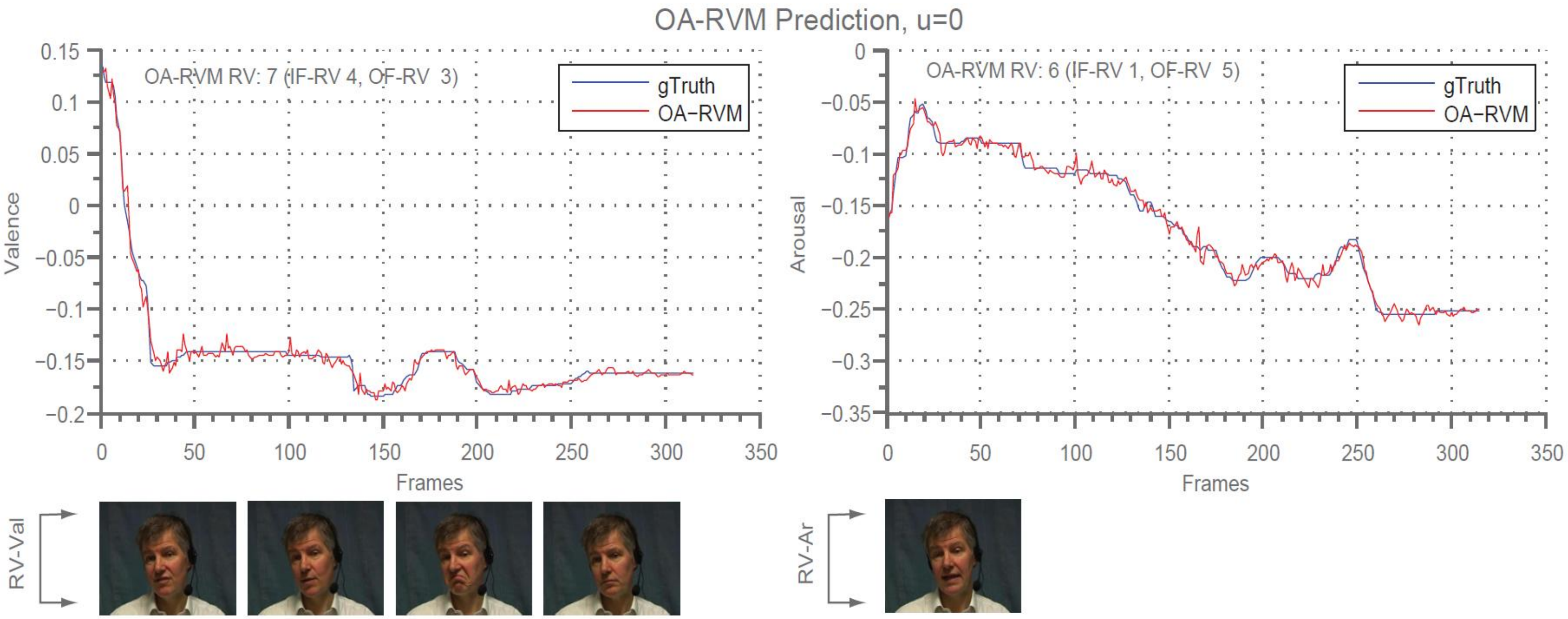
- RVM vs. OA-RVM:



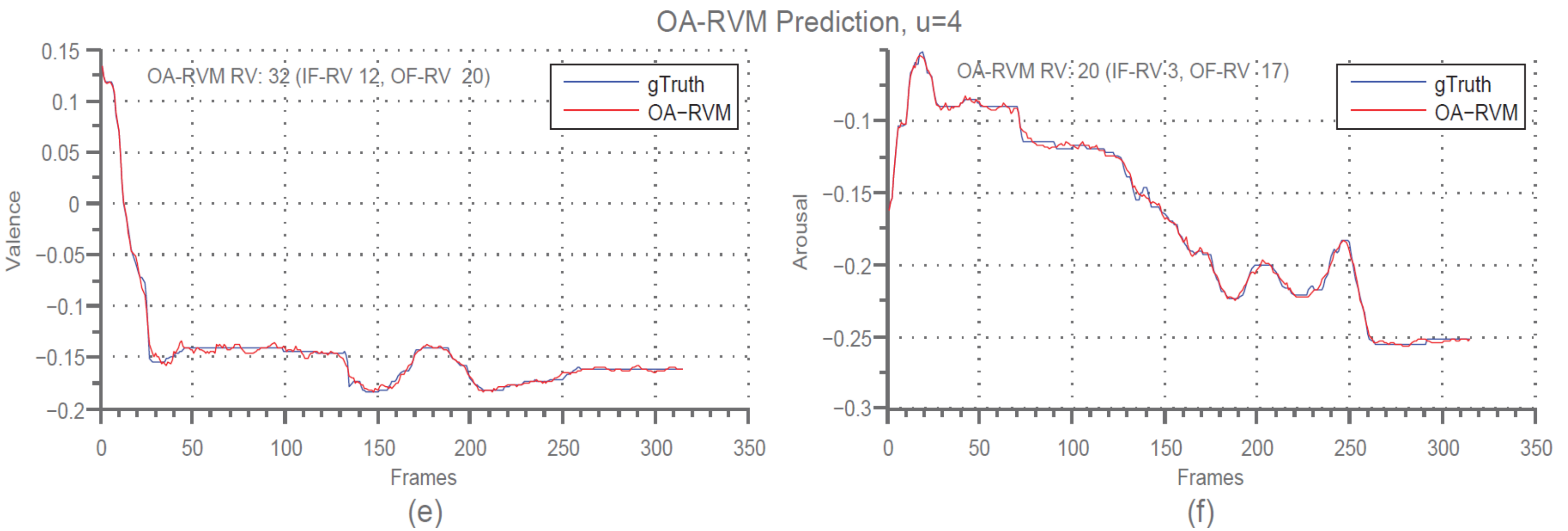


Institute for Anthropomatics, Computer Science Department

# Results: Prediction



# Results: Prediction



# Results: Prediction

- For both dimension (valence and arousal) OA-RVM improves the prediction results in all cases

TABLE III
SUBJECT-DEPENDENT PREDICTION RESULTS (RMSE).

| | Valence | | | Arousal | | |
|---|---|---|---|---|---|---|
| **POS** | **RVM** | **RVM-OA** | **v** | **RVM** | **RVM-OA** | **v** |
| subj1 | 0.16 | 0.15 | 10 | 0.13 | 0.11 | 10 |
| subj2 | 0.17 | 0.13 | 18 | 0.14 | 0.13 | 5 |
| subj3 | 0.11 | 0.09 | 12 | 0.10 | 0.09 | 18 |
| subj4 | 0.17 | 0.15 | 8 | 0.23 | 0.19 | 18 |
| **NEG** | **RVM** | **RVM-OA** | **v** | **RVM** | **RVM-OA** | **v** |
| subj1 | 0.14 | 0.10 | 12 | 0.30 | 0.29 | 14 |
| subj2 | 0.11 | 0.09 | 18 | 0.37 | 0.33 | 9 |
| subj3 | 0.08 | 0.07 | 18 | 0.22 | 0.21 | 18 |
| subj4 | 0.11 | 0.10 | 18 | 0.48 | 0.40 | 12 |

TABLE IV
SUBJECT-INDEPENDENT PREDICTION RESULTS (RMSE)

| | Valence | | | | Arousal | | | |
|---|---|---|---|---|---|---|---|---|
| **POS** | **SVR** | **RVM** | **RVM-OA** | **v** | **SVR** | **RVM** | **RVM-OA** | **v** |
| subj1 | 0.21 | 0.16 | 0.15 | 18 | 0.16 | 0.16 | 0.15 | 18 |
| subj2 | 0.22 | 0.26 | 0.17 | 18 | 0.18 | 0.18 | 0.14 | 9 |
| subj3 | 0.22 | 0.22 | 0.22 | 12 | 0.17 | 0.17 | 0.16 | 12 |
| subj4 | 0.19 | 0.16 | 0.15 | 6 | 0.19 | 0.14 | 0.13 | 18 |
| **NEG** | **SVR** | **RVM** | **RVM-OA** | **v** | **SVR** | **RVM** | **RVM-OA** | **v** |
| subj1 | 0.11 | 0.10 | 0.09 | 12 | 0.36 | 0.39 | 0.35 | 18 |
| subj2 | 0.14 | 0.11 | 0.09 | 14 | 0.37 | 0.33 | 0.32 | 10 |
| subj3 | 0.10 | 0.10 | 0.10 | 5 | 0.37 | 0.40 | 0.37 | 18 |
| subj4 | 0.13 | 0.11 | 0.09 | 18 | 0.14 | 0.13 | 0.13 | 2 |

- In accordance with psychological evidence, arousal appears to be more challenging to model and predict, for the negative class
- Optimal window size appears to be subject and data-dependent.

# Conclusions

- OA-RVM augments RVM.
- OA-RVM outperforms both RVM and SVR:
  - Using a temporal (output) window.
  - Optimal temporal window may vary depending on the data at hand or the task.

- OA-RVM appears to provide a more sparse model.
- Future work should evaluate the propose model in a larger number of subjects.

# Discussion

- Refinement of emotional models
- Technical difficulties around the proposed models
- Data acquisition in unconstrained environments
- Baseline and ground truth recognition/agreement