UNIVERSITÉ DE GENÈVE
FACULTÉ DES SCIENCES
Département d'informatique

Estimating emotions and tracking interest during movie watching, based on multimedia content and physiological responses

Joep Kierkels, Mohammad Soleymani, Guillaume Chanel,
Eric Bruno, Stéphane Marchand-Maillet, Thierry Pun

CVML – CUI – University of Geneva

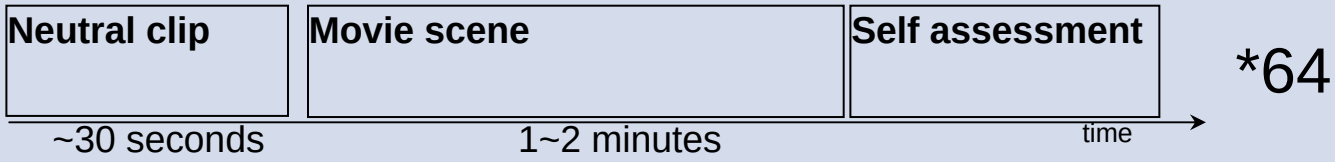
Why estimate emotions?

- Emotions reveal a persons true feelings in real-time
- In a video context, build a personal profile for video retrieval
- Physiological signals are (also real-time) related to these emotions

- Find correlated features
- Assessment of user behaviour and emotions
  - User's emotion prediction by arousal/valence levels from audio-video content analysis
  - User's emotional characterization by arousal/valence levels from physiological signals
- Assess interest levels

Video dataset

- In the experimental data set we have 8 movies from 4 different genres:
  - Horror: The Ring, 28 days later
  - Action: Kill Bill VOL I, Saving private Ryan
  - Drama: Hotel Rwanda, The Pianist
  - Comedy: Mr. Bean's Holiday, Love actually
- 8 short video clips were extracted from each movie
- A Neutral clip between each two video clips to record baseline and let the participant to return to neutral state
- Total duration ~2 hours



Multimedia content features

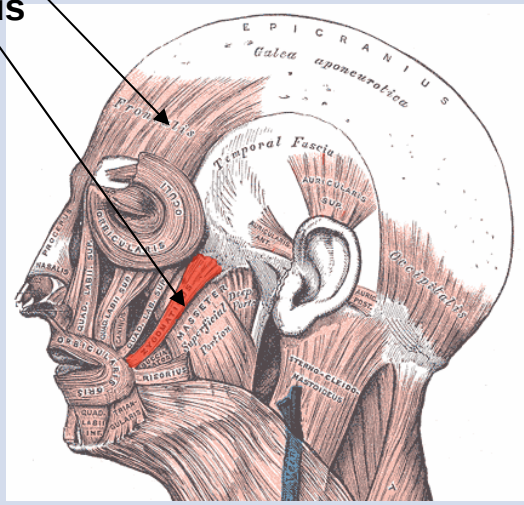
- All video content features are extracted with the help of OVAL and OMT software packages (*Viper*)
- Features extracted from multimedia content, namely:
  - Average shot duration
  - Shot change rate and variation
  - Color variance
  - Key lighting
  - Zero crossing rate (ZCR)
  - Audio energy
  - Audio type vector (music, speech, environment sound, and silence ratio over time)

Physiological feature extraction

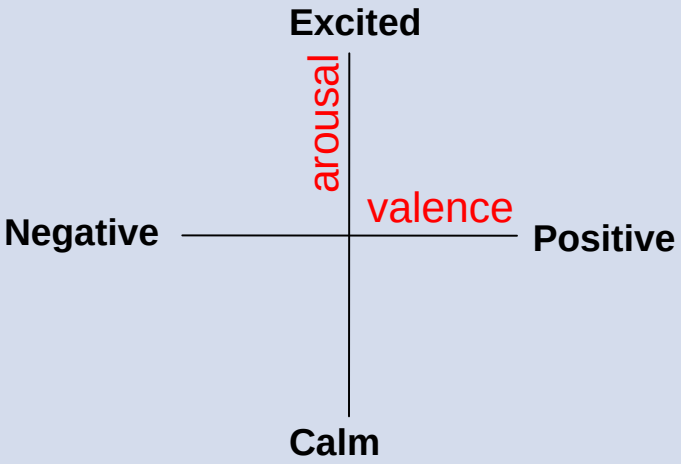
EMG	Power of Contraction
ECG	Heart Rate
	Heart Rate Variability
	Inter-Beat- Interval
Blinks	Blink Amplitude
	Blink Duration
	Blink Frequency
Plethysmograph	Heart Rate
	Heart Rate Variability
	Inter-Beat- Interval
Respiration	Respiration Depth
	Respiration Rate
GSR	GSR
	Number of peaks
	Peak amplitude
	dGSR/dt
Temperature	Temperature
	dTemp/dt

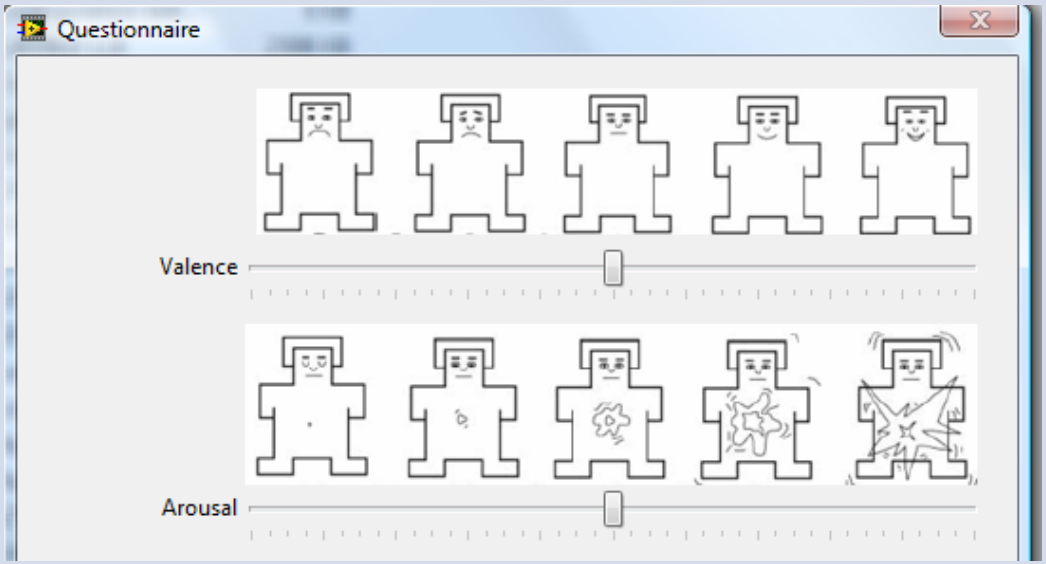
Frontalis

Zygomaticus



- Introducing the arousal valence space





Feature correlation

Correlation of selected multimedia features with physiological features of 8 participants				
	EMG Zygomatic. Energy/Key lighting	Skin temp.standard deviation /5th autocorrelation of MFCC coefficient	Skin temperature range/Shot length variation	EMG Zygomatic. energy/ 15th/20 bin of hue histogram
1	0.24	-	-	-0.41
2	0.62	0.44	0.42	-0.41
3	0.46	0.40	0.56	-0.34
4	0.40	0.32	0.43	-0.30
5	0.36	0.39	0.58	-
6	0.44	0.31	0.51	-0.32
7	0.47	0.34	0.27	-0.43
8	0.54	0.34	0.42	-0.45

Affect characterization

$$\begin{bmatrix} Y \end{bmatrix} = \begin{bmatrix} weights \end{bmatrix} \cdot \begin{bmatrix} feature - \\ values \end{bmatrix}$$

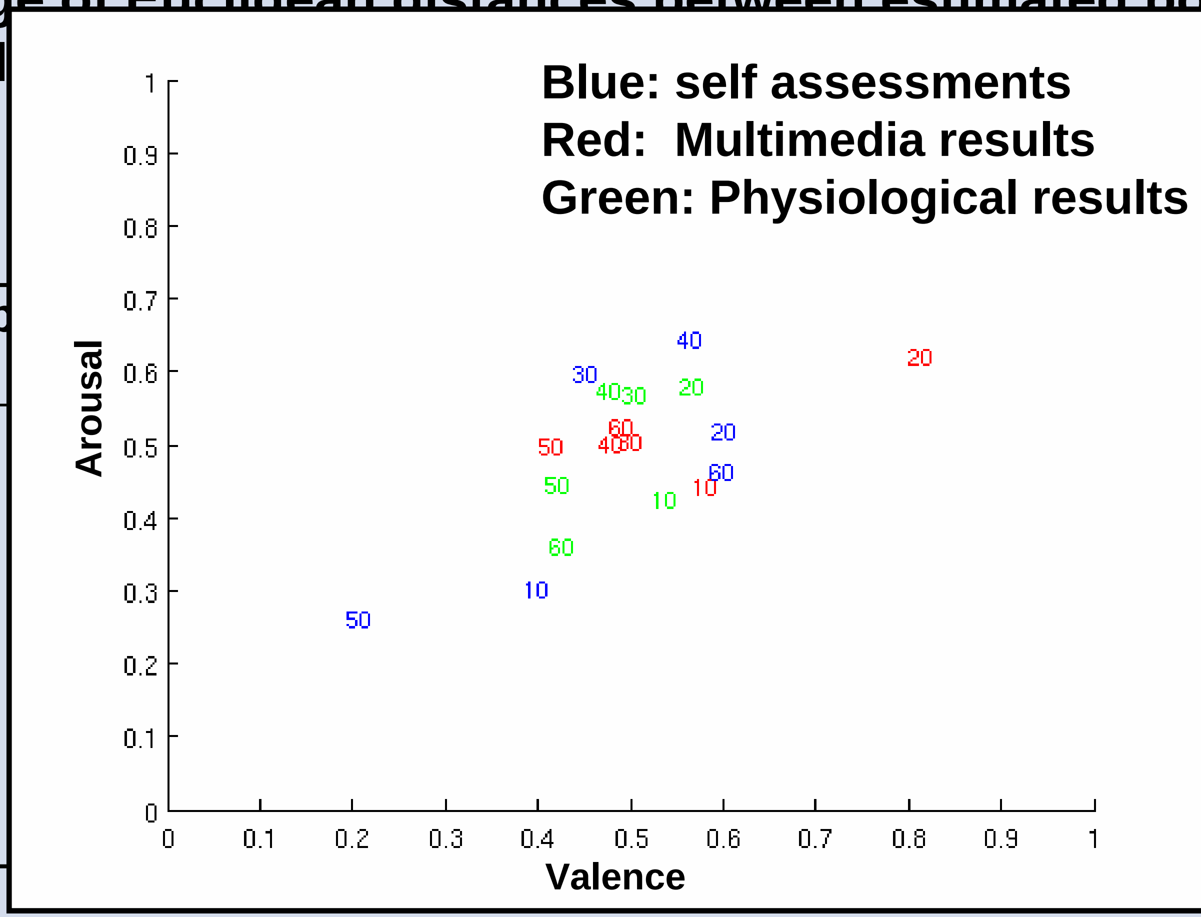
- *Y* can be either arousal or valence value, weights are trained. (RVM-lin.)
- Use computed weights to estimate arousal/valence scores (Tipping toolbox)
- Leave one out cross validation

$$\hat{y}(j) = \sum_i w_i x_i + w_0$$

i = feature number

j = movie scene number

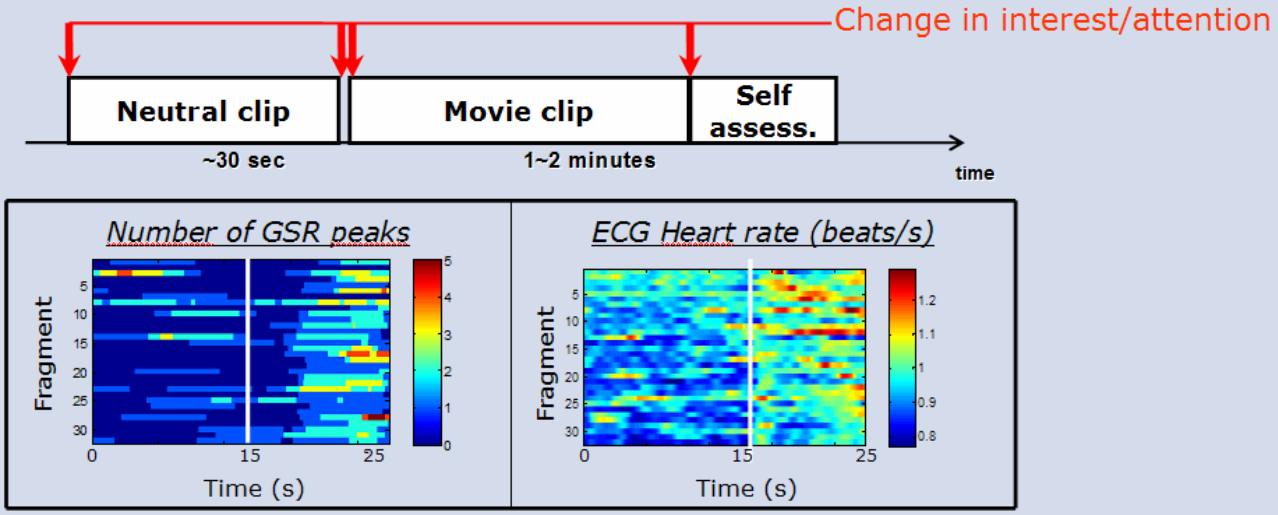
Average of Euclidean distances between estimated points and sel... for



Particip...	...nces with features
1	
2	
3	
4	
5	
6	
7	
8	

Interest detection

Need for training data containing relevant features for two levels of interest

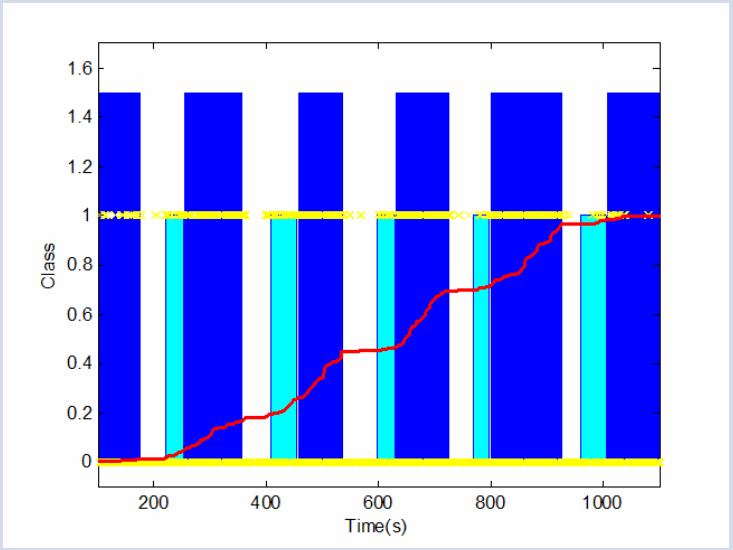


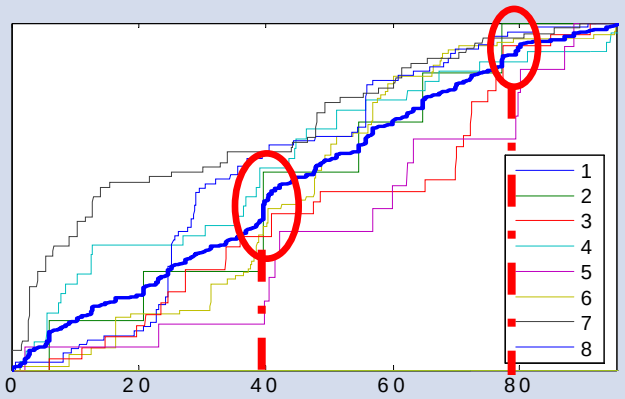
LDA Classifier output: 72% accuracy on training data

Interest detection during movies

- Analyze full experiment
- Compare participants
- Evaluate

00 : 00





Conclusions

- The correlation between stimuli (multimedia) and response (physiological signals) was shown.
- Results show the ability of the physiological signals for affective characterization.
- Temporal detection of interesting movie fragments is feasible.

Future work

- Using prior information for affect estimation (genre, users rating, etc)
- Modeling and estimating dynamics of affect in movie watching by multimedia features.
- Implement other features (e.g., head movements from camera).
- Obtain objective measure for detecting interesting episodes instead of visual inspection.