DECAF: MEG-based Multimodal Database for Decoding Affective Physiological Responses

Mojtaba Khomami Abadi[1,5]
Ramanathan Subramanian[2] , Seyed Mostafa Kia[1,3]
Ioannis Patras[4], Paolo Avesani[3]
Nicu Sebe[1]

[1]University of Trento
[2]Advanced Digital Sciences Center (ADSC), University of Illinois at Urbana-Champaign, Singapore
[3]NeuroInformatics Laboratory (NILab), Fondazione Bruno Kessler
[4]Queen Mary University of London
[5]Semantic, Knowledge and Innovation Lab (SKIL), Telecom Italia

FG 2015

Modalities in the State of the Art

Facial expressions (Posed, Spontaneous)

Physiological Signals

?

Amusement
Frustration
Sadness / Anger
Shock / Fear ….

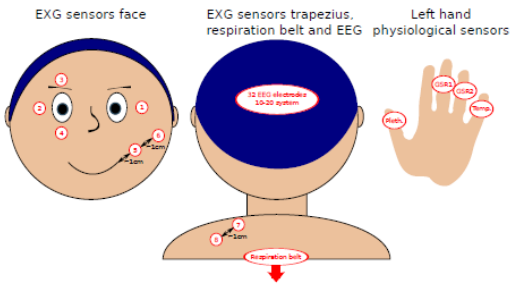
?

?

Brain Signals

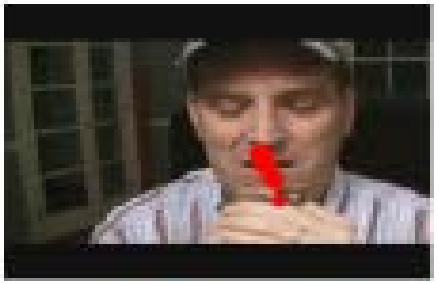
Behaviors (verbal, non-verbal)

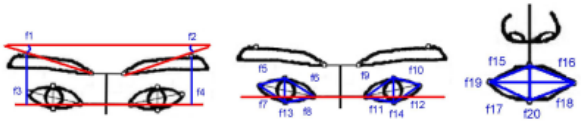
Existing datasets



EEG Brain Signals
Physiological Signals
Facial Videos
(Music Video Clips)

DEAP - Koelstra et al. [6], 2012





EEG Brain Signals
Physiological Signals
Eye Gaze
Facial Videos
(Movie Clips - Images)

MAHNOB- Soleymani et al. [7], 2012

Motivation

- Limitations with sensing equipment mean evaluations using stimuli presented over a short time interval and small user population.
- MEG is less invasive and the user response is less affected by psychological stressors enabling naturalistic experience.
- Suitability of stimuli – Movie clips vs. Music videos

DECAF:

- a dataset comprising concurrent recording of:
  MEG signals
  Physiological signals
  NIR Facial Video

  } 30 Subjects x (40 Music + 36 Movie Clips)

  while the subjects look at emotional video clips.

EEG for emotion recognition:

Koelstra et al. [6]: EEG effectively encodes the emotional responses of human to music videos

Our assumption: MEG also encodes emotional responses

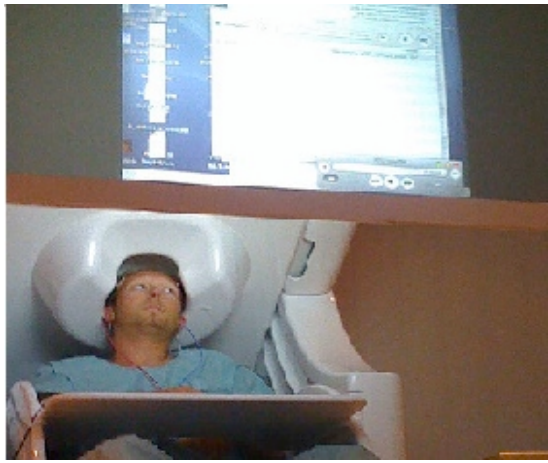
EEG

MEG vs EEG:

- Both EEG and MEG have high temporal resolution (1-5kHz)

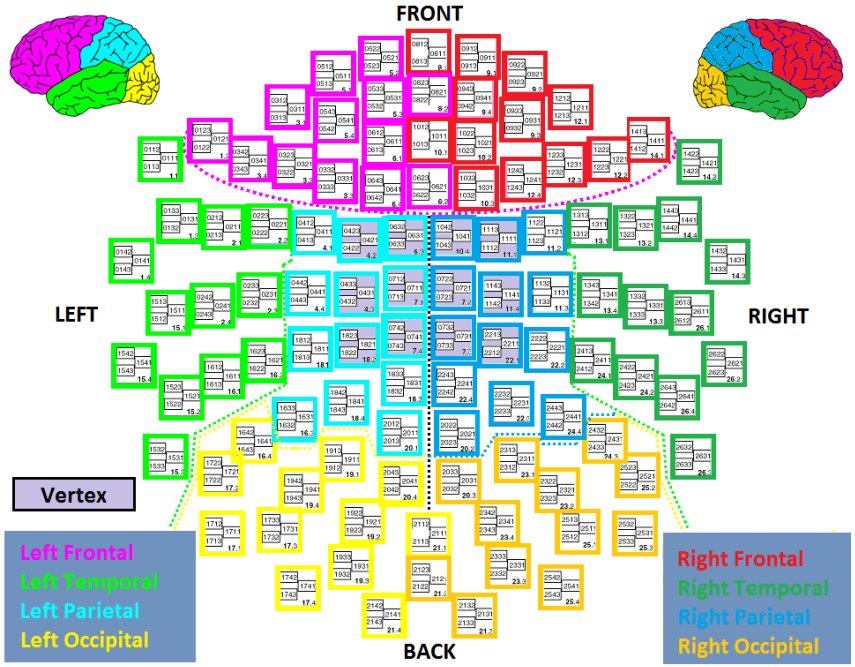
- EEG sensor is much cheaper in price than MEG
- EEG is somewhat invasive
- EEG signal is distorted by hair and scalp bone

- MEG is much more accurate and gives a very good spatial resolution of brain responses (#channels: 306 > 32)
- Brain activity source localization techniques work well for MEG but are not suitable with EEG
- MEG is less invasive than EEG (no direct contact)



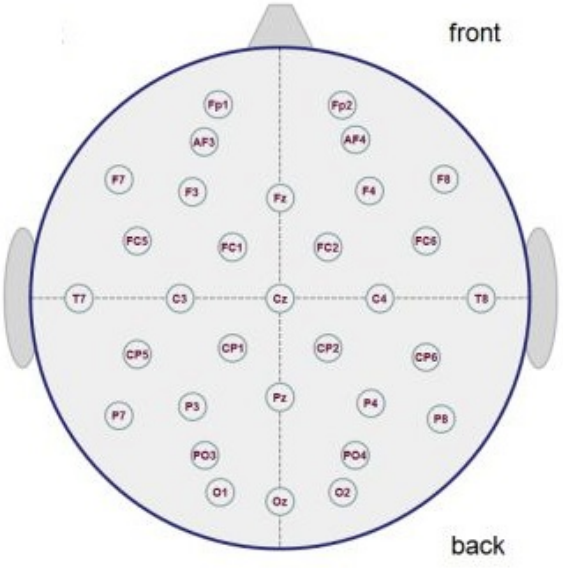
MEG

MEG vs EEG



306 Channels

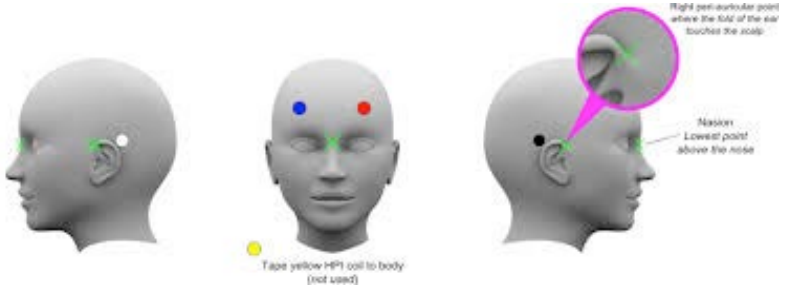
Elekta NeuroMag Device



32 Channels

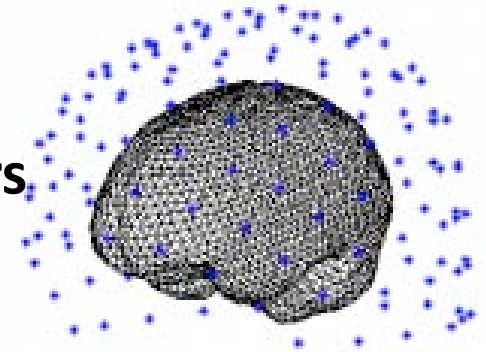
Bio-semi EEG Device

DECAF MEG includes:

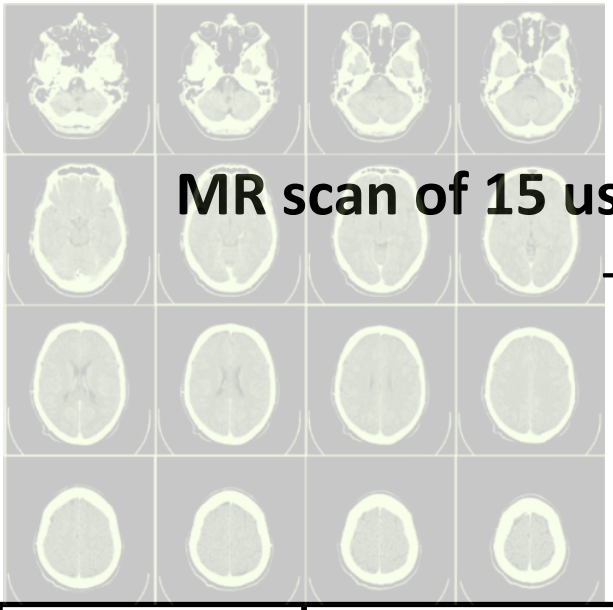


Time-Continuous recordings of Head Pose Indicators (HPIs)

- a 300 Hz signal to locate the exact pose of head under MEG

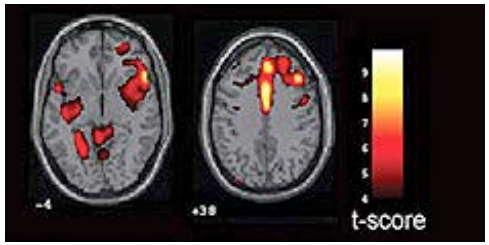


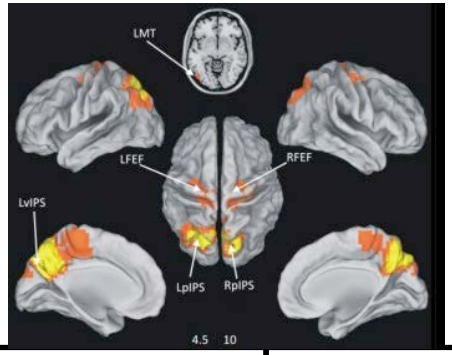
3D Scan of the users' head shape for all the users



MR scan of 15 users for source-brain activity analysis

- will become available soon

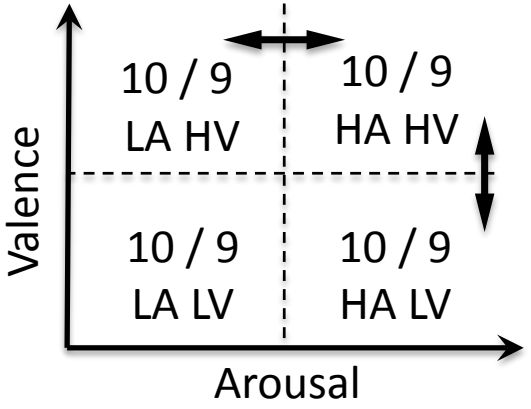




Stimuli :

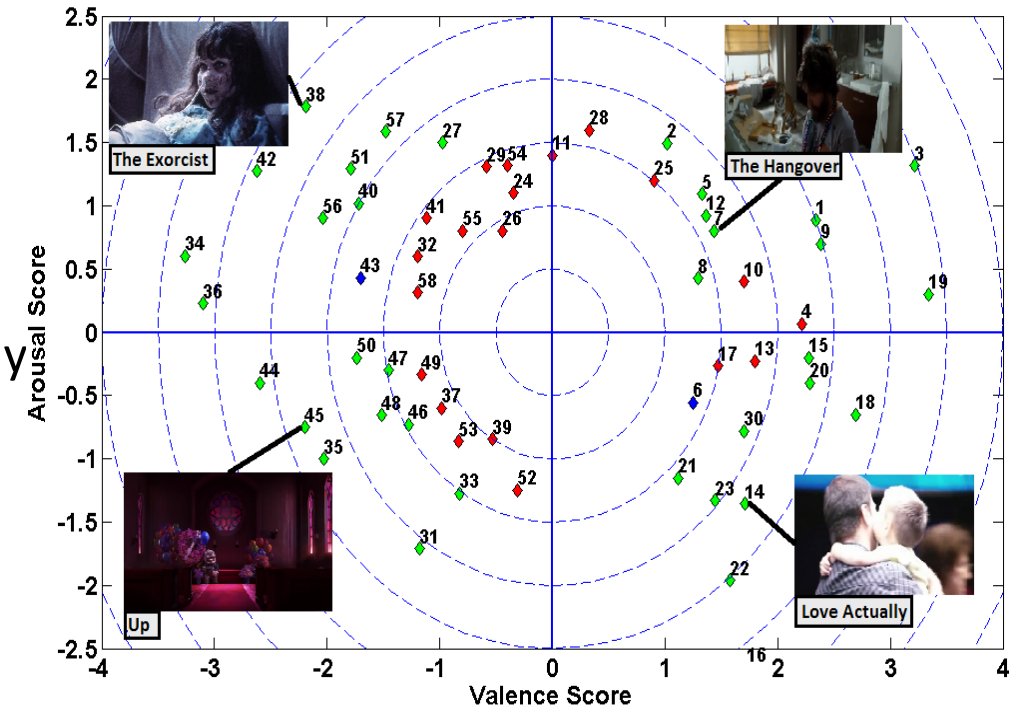
Music video clips:

- 40 music video clips proposed by DEAP [6]
- 60 second highlight for each music video clip



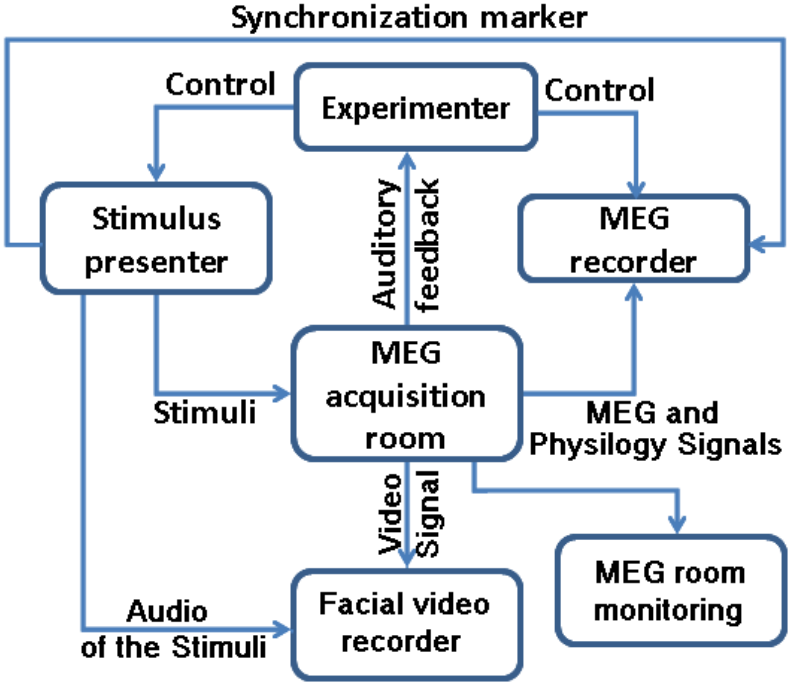
Movie videos:

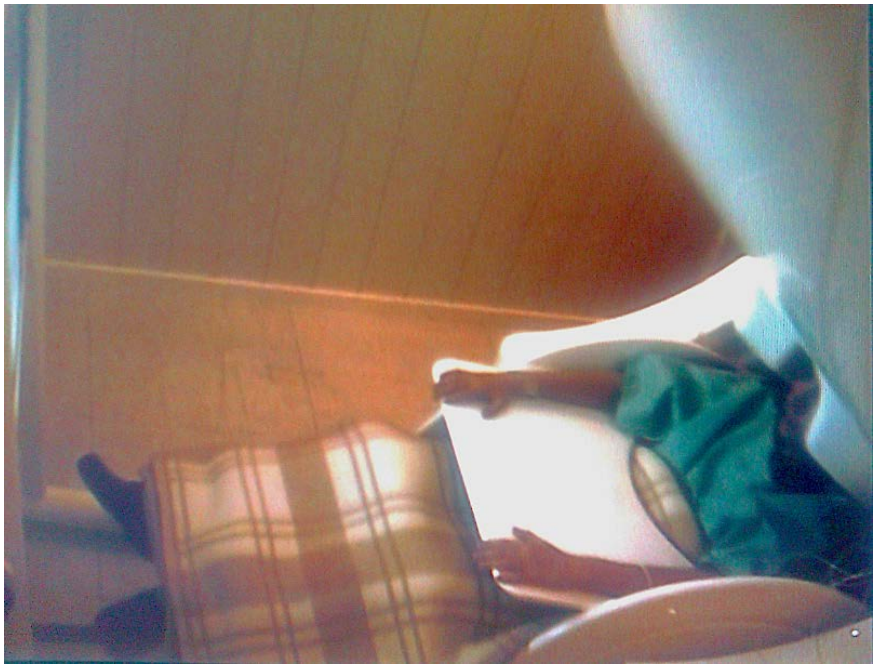
- 36 movie clips
- varying length (51” to 128”)
  Mean=80” Std. = 20”
- Selected based on a preliminary study with a pool of 58 clips and 42 graduate students



Experiment setup

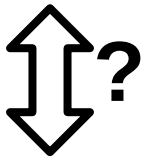
- 30 subjects (16 male, 14 female, age range 27.3 ± 4.3)
- 76 stimuli (40 music video clips + 36 movie video clips)
- MEG brain signal , ECG , horizontal EOG (hEOG), Trapezius EMG (tEMG), NIR facial videos recorded
- Subjects' self-assessments



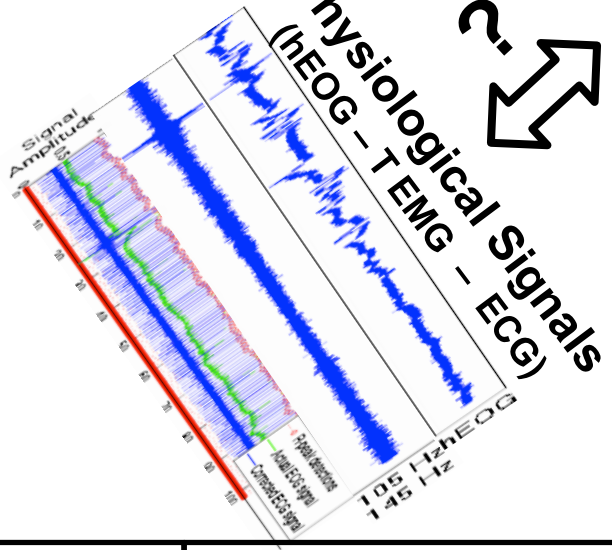


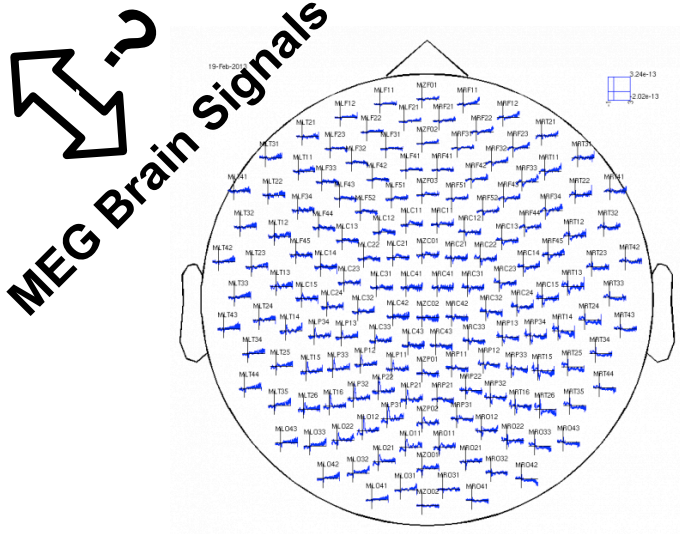
Modalities

Facial videos



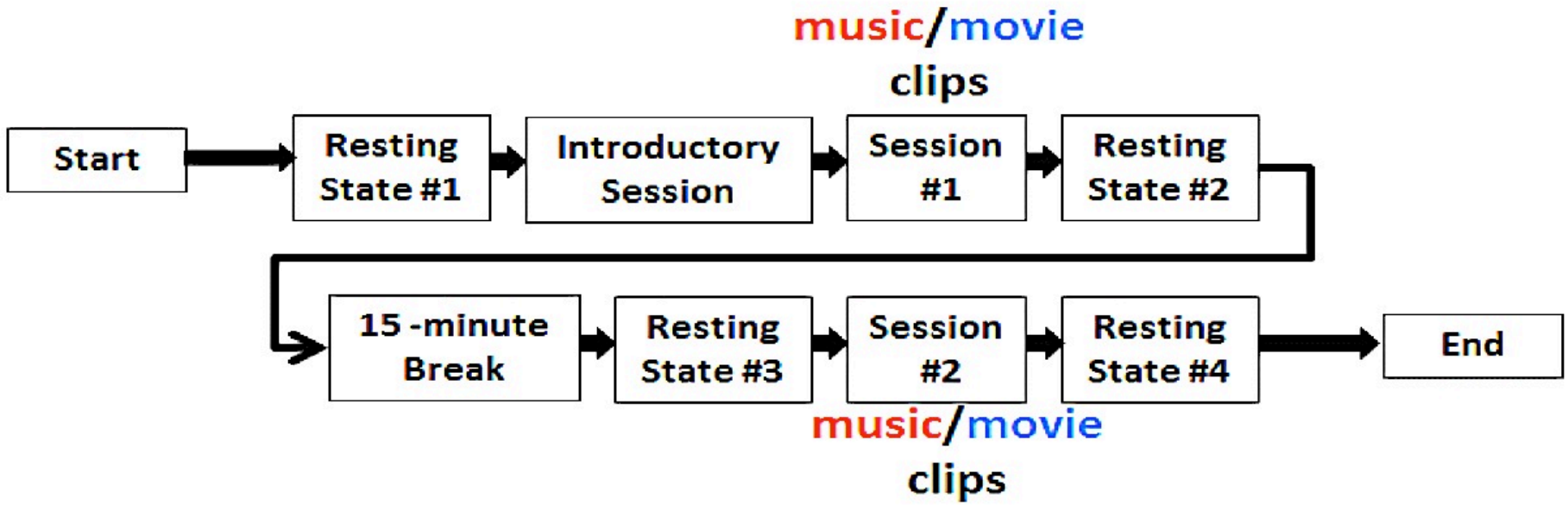
Arousal
Valence
Dominance
Familiarity



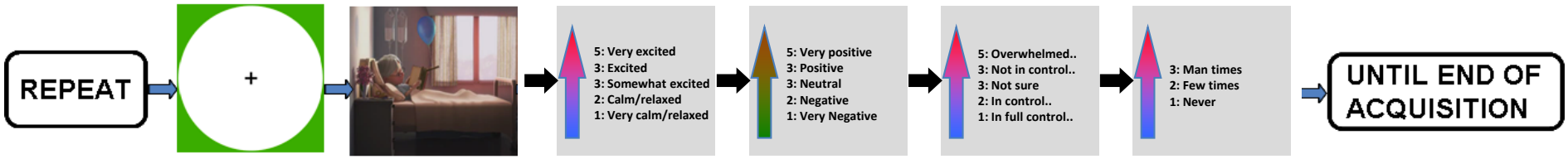


Experiment setup – Protocol :

MATLAB – PsychToolbox + ASF



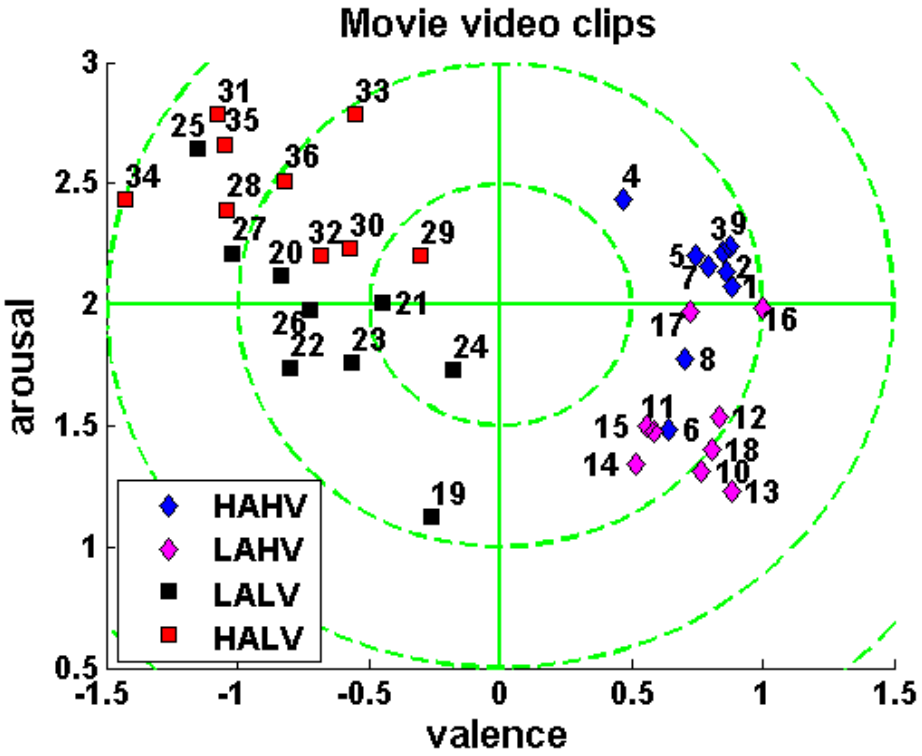
Structure of each Session :

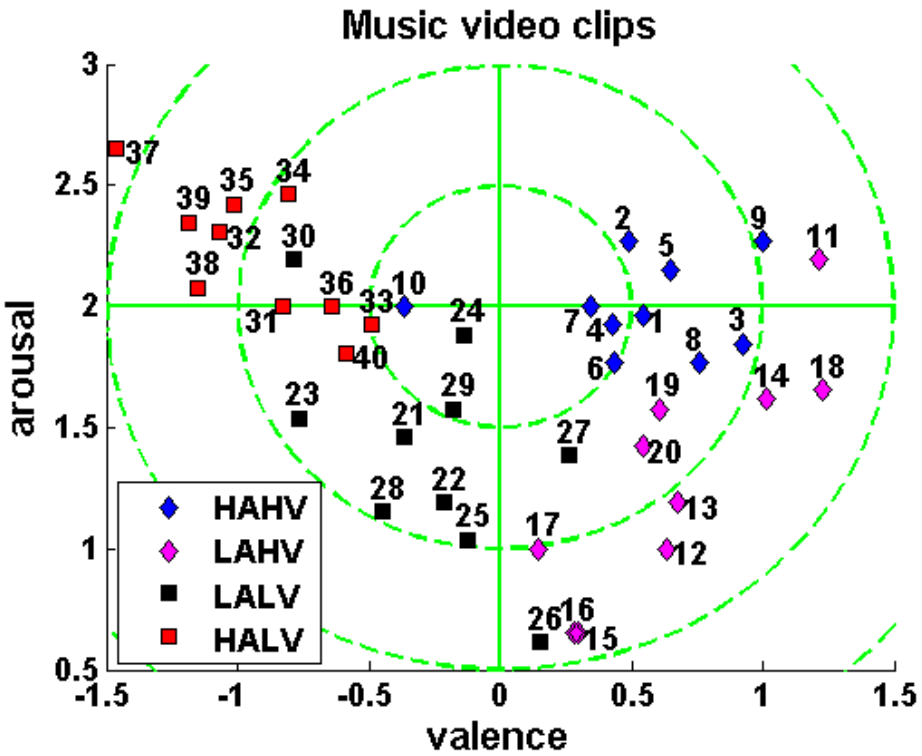


SHOW FIXATION (4 Secs) | PRESENT MOVIE/MUSIC STIMULUS | AROUSAL RATING (Max 15 Secs) | VALENCE RATING (Max 15 Secs) | DOMINANCE RATING (Max 15 Secs) | FAMILIARITY RATING (Max 15 Secs)

SINGLE TRIAL

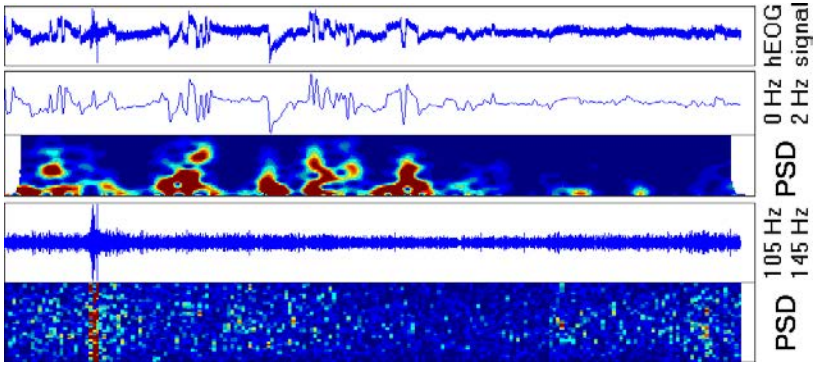
Distribution of self-assessment ratings:





Psycho-Physiological signal analysis

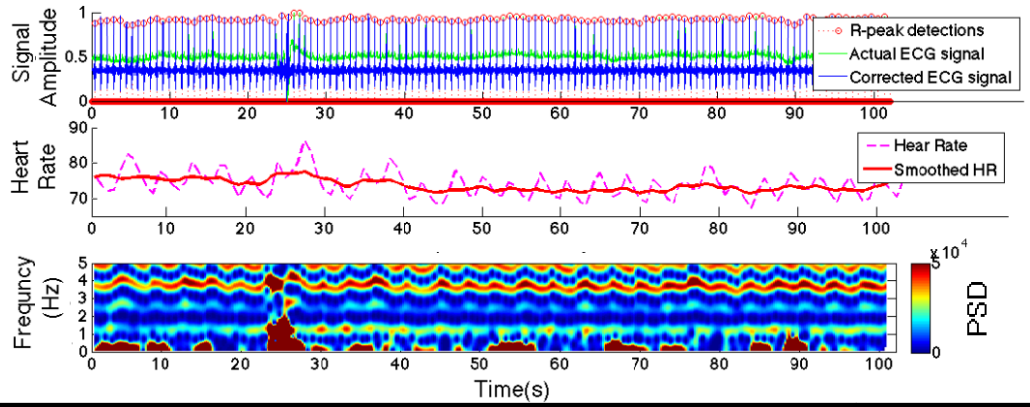
Facial Landmark Tracks



Horizontal Electrooculography

Trapezius Electromyography

Electrocardiography



MEG Correlates with emotions

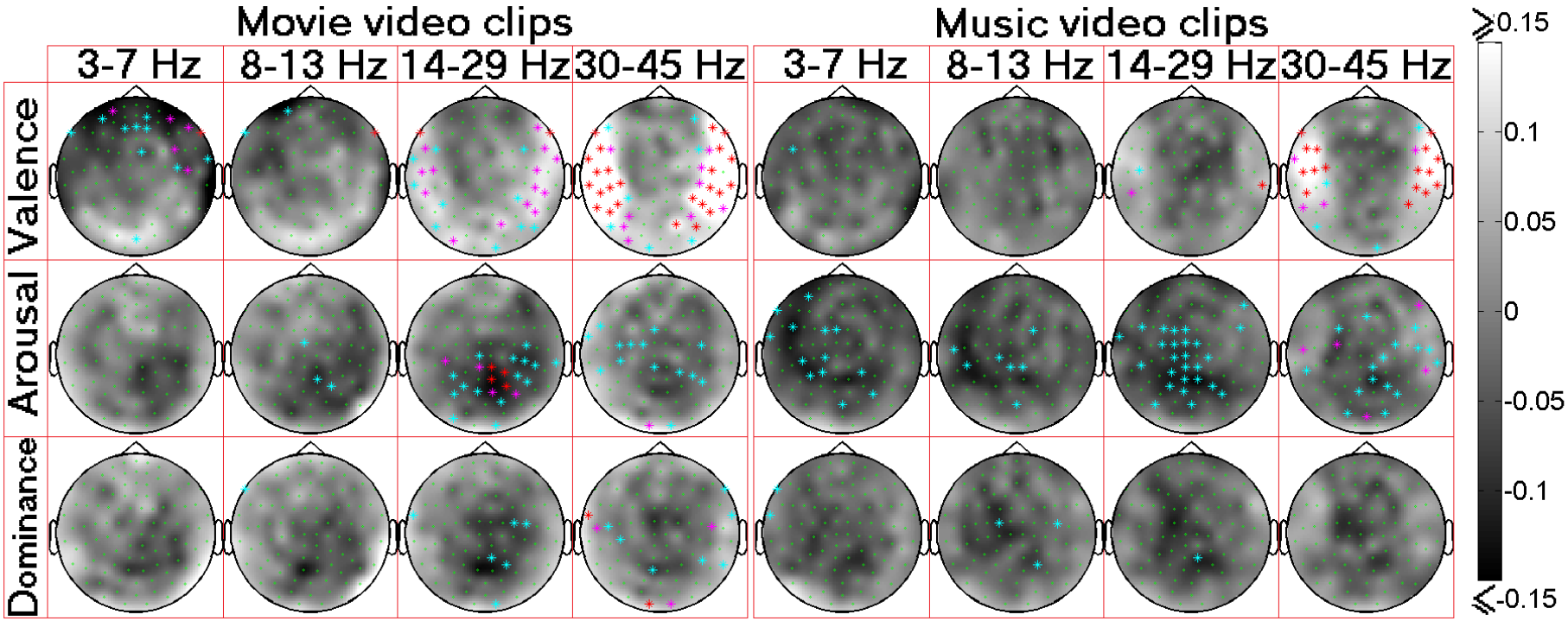


Fig. 7. Spearman correlation analysis between the MEG responses and participants' self-assessments. Correlation over each channel (in green) is denoted by the gray level, and significant ($p < 0.05$, $p < 0.01$, and $p < 0.001$) correlations are highlighted with $*$ marks (in cyan, magenta, and red).

Classification Results using Linear SVM:

$^{*} = p < 0.05, \; ^{**} = p < 0.01, \; ^{***} = p < 0.001$

		Movie (PB)			Music (PB)			Movie (SB)			Music (SB)		
		A	V	D	A	V	D	A	V	D	A	V	D
Vertex	Acc	0.59	0.57	0.57	0.51	0.51	0.52	0.55	0.55	0.51	0.53	0.50	0.53
	F1	0.58***	0.57***	0.57***	0.51	0.51	0.51	0.54	0.53	0.48	0.52	0.49	0.49
Left Temporal	Acc	0.60	0.60	0.58	0.51	0.51	0.52	0.59	0.58	0.51	0.54	0.50	0.54
	F1	0.60***	0.60***	0.58***	0.51	0.51	0.51	0.59***	0.57**	0.49	0.52	0.49	0.51
Right Temporal	Acc	0.62	0.56	0.57	0.55	0.53	0.53	0.59	0.55	0.54	0.60	0.54	0.54
	F1	0.62***	0.55**	0.57***	0.55*	0.53*	0.53*	0.58**	0.53	0.51	0.58***	0.53	0.51
Left Parietal	Acc	0.60	0.56	0.57	0.52	0.52	0.55	0.55	0.56	0.53	0.53	0.48	0.52
	F1	0.60***	0.55**	0.57***	0.52	0.51	0.54*	0.54*	0.54*	0.49	0.52	0.47	0.49
Right Parietal	Acc	0.58	0.57	0.57	0.51	0.51	0.52	0.55	0.55	0.58	0.51	0.53	0.54
	F1	0.57**	0.57***	0.56***	0.50	0.50	0.52	0.53	0.53	0.55**	0.50	0.52	0.51
Left Occipital	Acc	0.58	0.59	0.57	0.51	0.50	0.52	0.53	0.56	0.54	0.55	0.48	0.53
	F1	0.57**	0.58***	0.56**	0.51	0.50	0.52	0.51	0.54*	0.50	0.54*	0.47	0.50
Right Occipital	Acc	0.60	0.56	0.56	0.50	0.53	0.50	0.57	0.54	0.55	0.54	0.53	0.53
	F1	0.60***	0.55**	0.56*	0.50	0.53	0.50	0.56**	0.53	0.52	0.53	0.51	0.49
Left Frontal	Acc	0.59	0.56	0.57	0.55	0.51	0.51	0.56	0.56	0.53	0.57	0.55	0.60
	F1	0.58***	0.56***	0.57***	0.54*	0.50	0.51	0.55**	0.55**	0.50	0.55**	0.54*	0.56**
Right Frontal	Acc	0.55	0.59	0.61	0.50	0.52	0.50	0.51	0.54	0.53	0.54	0.52	0.53
	F1	0.55***	0.59***	0.61***	0.49	0.52	0.49	0.50	0.53	0.49	0.53	0.51	0.49
MEG Early Fusion	Acc	0.60	0.61	0.59	0.53	0.53	0.54	0.55	0.58	0.55	0.58	0.56	0.55
	F1	0.60***	0.61***	0.59***	0.52	0.53	0.54*	0.54*	0.58***	0.53	0.55**	0.55**	0.53*
Peripheral Physiology	Acc	0.55	0.60	0.50	0.55	0.59	0.56	0.56	0.60	0.56	0.57	0.55	0.57
	F1	0.54*	0.59***	0.50	0.54*	0.59***	0.55**	0.55**	0.59***	0.54*	0.56**	0.54*	0.54**
Facial Expressions	Acc	0.58	0.64	0.53	0.60	0.61	0.53	0.56	0.61	0.55	0.58	0.60	0.55
	F1	0.57**	0.64***	0.53	0.59**	0.60***	0.53	0.54**	0.61***	0.54	0.56**	0.58***	0.52
Multimedia Content	Acc	0.58	0.64	0.33	0.85	0.73	0.57	0.52	0.61	0.53	0.62	0.68	0.58
	F1	0.57	0.64	0.33	0.85	0.72	0.57	0.51	0.60***	0.52	0.61***	0.67***	0.55*
Late Fusion	Acc	0.70	0.79	0.66	0.85	0.82	0.66	0.66	0.73	0.72	0.73	0.76	0.74
	F1	0.68***	0.77***	0.64***	0.84***	0.81***	0.65***	0.62***	0.71***	0.66***	0.70***	0.73***	0.67***

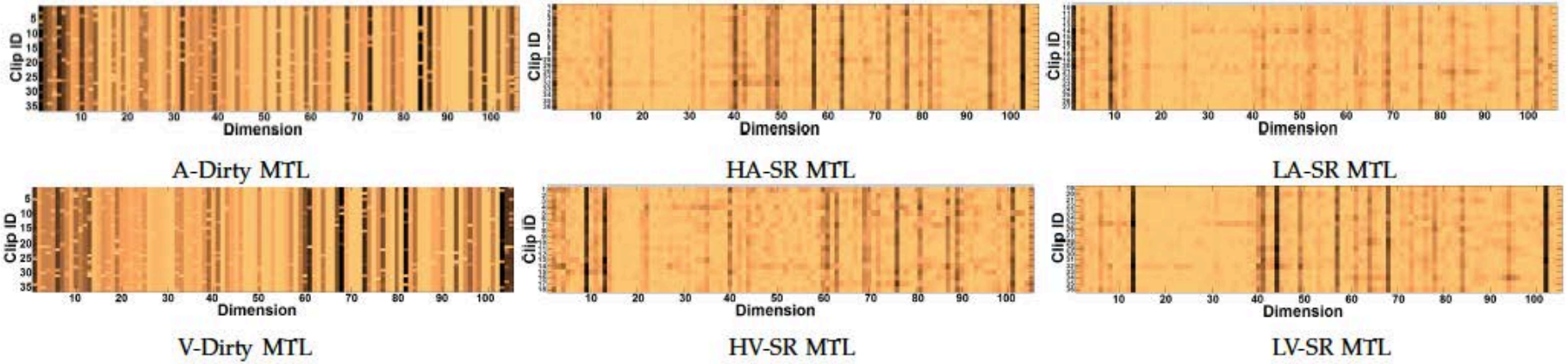
Continuous Emotion Estimation using Multi-task learning

			First		Second	
			5 s	15 s	5 s	15 s
Val	MM	Lasso	1.98±1.25	3.07±1.48	1.68±0.18	2.81±0.97
		MT-Lasso	**1.00±0.05**	1.66±0.54	**1.18±0.14**	**2.03±0.71**
		Dirty MTL	1.11±0.06	1.79±0.55	1.27±0.16	2.10±0.69
		SR MTL	1.09±0.09	**1.55±0.39**	1.89±0.13	2.80±0.74
	MEG	Lasso	1.30±0.09	1.87±0.46	2.03±0.25	2.93±0.78
		MT-Lasso	1.32±0.09	1.98±0.54	1.54±0.21	2.47±0.81
		Dirty MTL	1.42±0.10	2.44±0.82	**1.51±0.19**	**2.44±0.82**
		SR MTL	**1.09±0.05**	**1.58±0.41**	2.07±0.17	2.84±0.69
Asl	MM	Lasso	1.54±0.47	2.11±0.77	2.18±0.58	3.28±2.17
		MT-Lasso	**0.91±0.11**	1.47±0.47	**1.10±0.08**	1.89±0.66
		Dirty MTL	1.07±0.09	1.62±0.46	1.23±0.08	**1.97±0.61**
		SR MTL	1.01±0.07	**1.42±0.35**	1.86±0.13	2.48±0.53
	MEG	Lasso	1.11±0.08	1.65±0.45	1.75±0.06	2.53±0.66
		MT-Lasso	1.12±0.09	1.71±0.51	1.41±0.11	2.27±0.73
		Dirty MTL	1.19±0.11	1.84±0.56	**1.38±0.11**	**2.25±0.75**
		SR MTL	**0.99±0.08**	**1.42±0.36**	1.73±0.06	2.44±0.60

Valence/Arousal prediction with multimedia (MM) and MEG features.

RMSE mean, standard deviation over four runs are reported.

Range of V, A levels is [-3, 3]. Best model is shown in bold.



Learned weights for arousal (top) and valence (bottom) for the movie clips with Dirty MTL and SR MTL. Audio-visual features over the entire clip length were used for model training. Larger weights are denoted using darker shades.

Conclusion

- **Largest multi-modal affective dataset** with MEG, peripheral and facial responses: 30 subjects X (40 music video + 36 movie clips)

  over 2K samples

- Suitable for **cognitive-science natural-stimuli studies** with
  - time-continuous HPI
  - recorded head-shape-models
  - recordings of artifacts: ECG + tEMG + hEOG (includes facial EMG)
  - MR scans of 15 users for source brain activity analysis

- Includes time-continuous valence and arousal labels for the 36 movie video clips for

  time-continuous Multi-Media affective tagging

Thank you !

http://disi.unitn.it/~mhug/DECAF.html