Note: This is a read-only printable file. For complete presentation with multimedia, please contact me.

Utilizing Multi-modal Bio-sensing Toward Affective Computing in Real-world Scenarios

May 20, 2020

Ph.D. Final Defense

Siddharth
Ph.D. Candidate in Electrical Engineering
(Intelligent Systems, Robotics, and Control)
UC San Diego

Doctoral Committee Members

Professor Mohan M. Trivedi (Chair)
Professor Tzyy-Ping Jung (Co-Chair)
Professor Terrence J. Sejnowski
Professor Vikash Gilja
Professor Patrick P. Mercier

Five Ws and One H

- Who
- Where
- What
- Why
- When
- How

Five Ws and One H

- **Who** – Siddharth and collaborators
- **Where** – UC San Diego and Facebook Reality Labs
- **What**
- **Why**
- **When**
- **How**

Five Ws and One H

- **Who** – Siddharth and collaborators
- **Where** – UC San Diego and Facebook Reality Labs
- **What** is **Affective Computing**?
- **Why** use **Bio-sensing**?
- **When** are **Multi-modal** systems advantageous?
- **How** to apply them toward **Real-world** applications?

Five Ws and One H

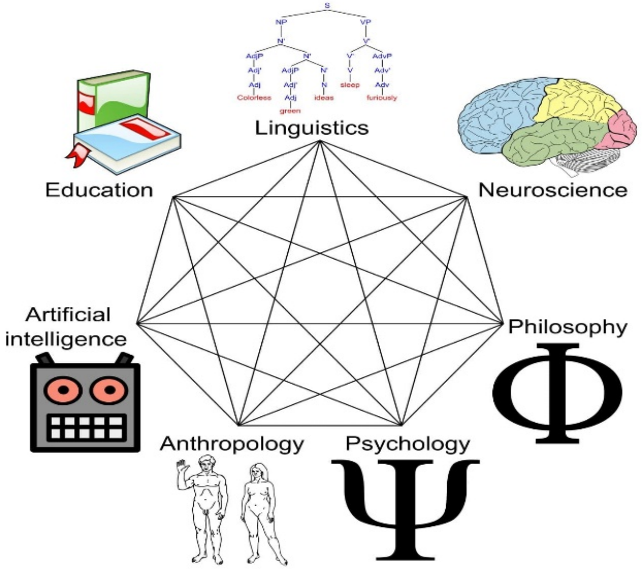
- **Who** – Siddharth and collaborators
- **Where** – UC San Diego and Facebook Reality Labs
- **What** is **Affective Computing**?
- **Why** use **Bio-sensing**?
- **When** are **Multi-modal** systems advantageous?
- **How** to apply them toward **Real-world** applications?

What is Affective Computing?

Affective Computing is the study and development of systems and devices that can recognize, interpret, process, and simulate human affects (feeling, emotion, or mood).[1]

Computer Science

Psychology



Cognitive Science

Affective Computing is a newer research field as compared to the study of emotions.

[1] Tao et al., Affective Computing: A Review, *International Conference on Affective computing and intelligent interaction*, 2005.

EMOTIONS

Probably as long as humans have been **self-aware**, they have wondered about the origin, essence, and utility of **emotions.**

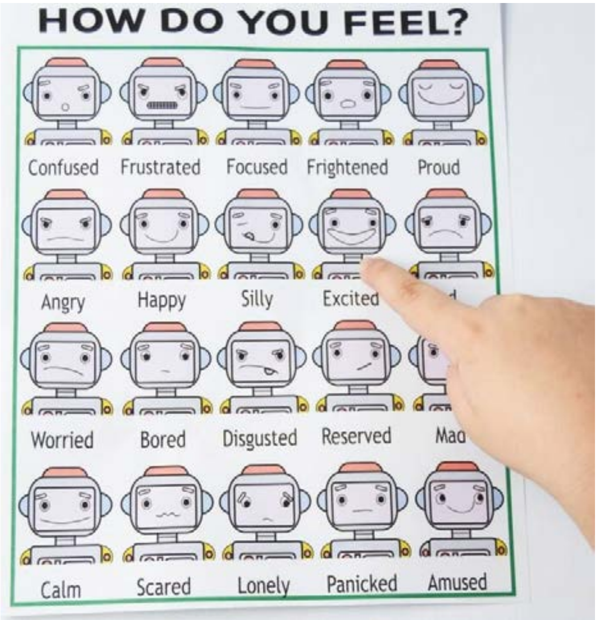
In **Western** (especially **Greek**) philosophy, emotions (**émouvoir**) were considered as playing a **destructive** role in decision-making.

In **Eastern** (especially **Buddhist**) philosophy, emotions (**bhāva**) were considered as a **hindrance** preventing liberation from suffering.

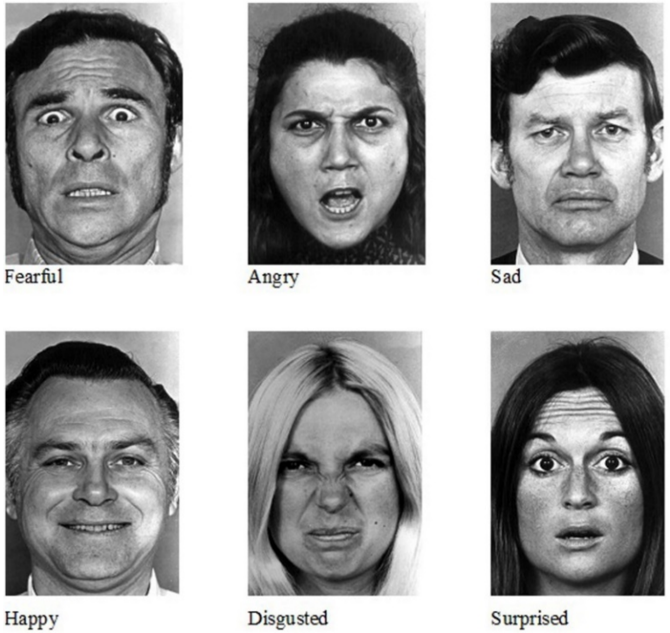
Tuske, J., The concept of emotion in classical Indian philosophy, *The Stanford Encyclopedia of Philosophy,* 2011

EMOTIONS

Such an **obsession** with emotions has naturally led to much research in studying their **origins** and **classifying** them into various categories. For centuries, **two methods** have been predominantly used to this end.



Receiving Human Feedback

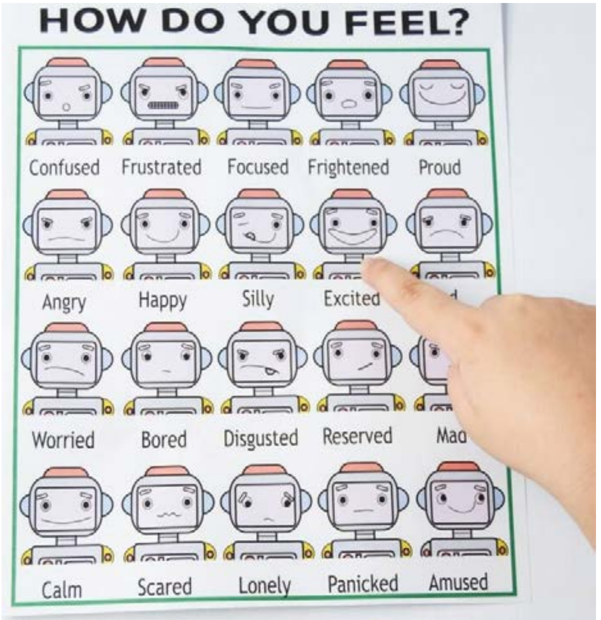


Recognizing Facial Expressions

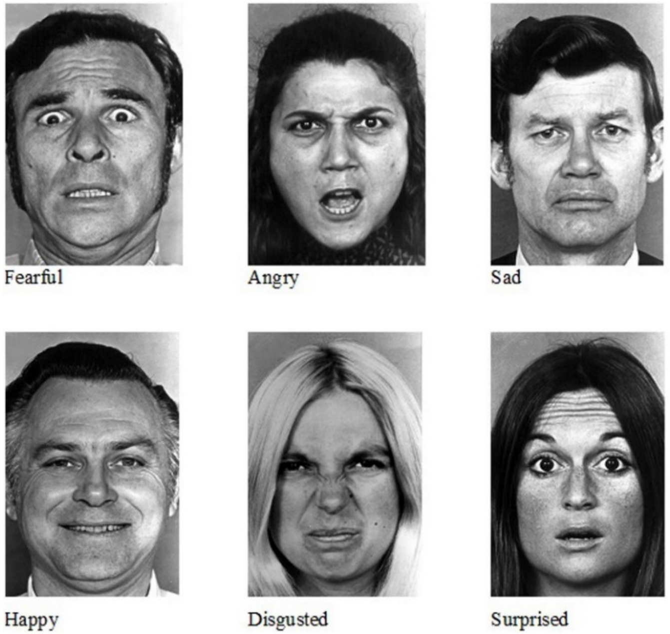
With developments in **electronics** and **computing** in the past half-century, it has now become possible for the **first time** in human history to utilize these **two methods** in an **automated** manner.

EMOTIONS

Such an **obsession** with emotions has naturally led to much research in studying their **origins** and **classifying** them into various categories. For centuries, **two methods** have been predominantly used to this end.



Receiving Human Feedback



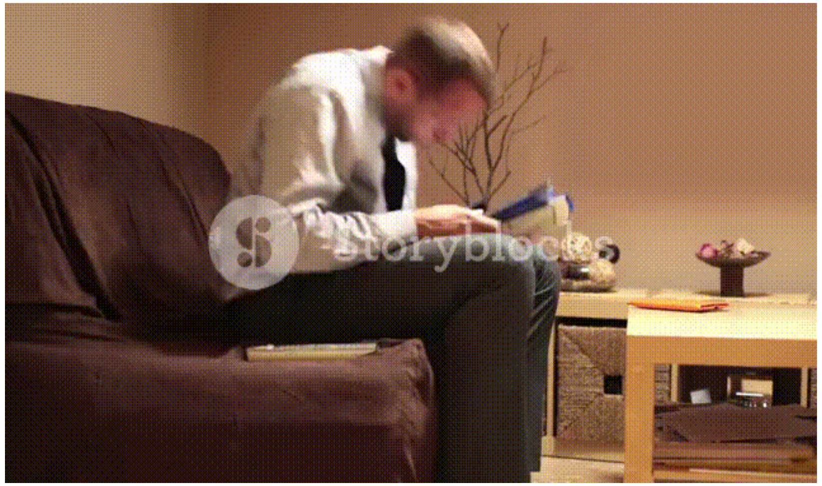
Recognizing Facial Expressions

These developments have emerged as a significant component of **Affective Computing.** However, the above **two methods** can be easily implemented in a system by a **joystick** and a **camera** respectively.

Five Ws and One H

- **Who** – Siddharth and collaborators
- **Where** – UC San Diego and Facebook Reality Labs
- **What** is **Affective Computing**?
- **Why** use **Bio-sensing**?
- **When** are **Multi-modal** systems advantageous?
- **How** to apply them toward **Real-world** applications?

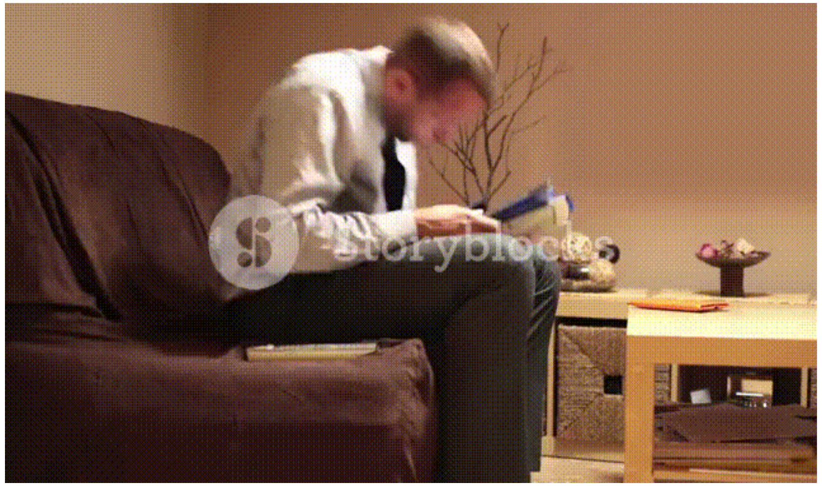
Why use Bio-sensing?



Intelligent Assistant: Hmmmm….
I detect that you are upset. Here, this should help.

(Plays your favorite song and turns on the television.)

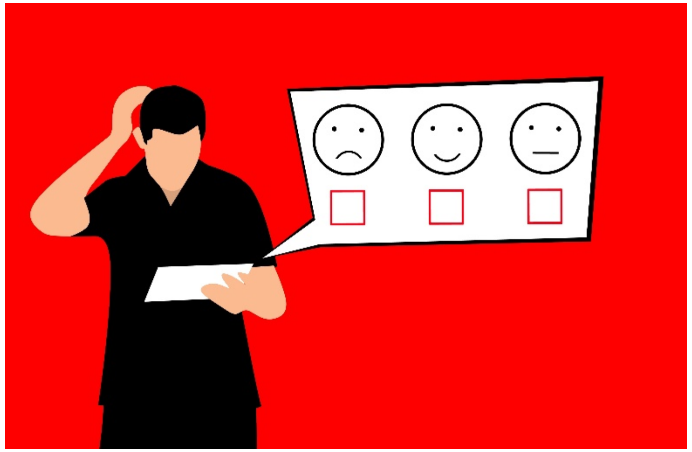
Why use Bio-sensing?



Intelligent Assistant: Hmmmm…. I detect that you are upset. Here, this should help.

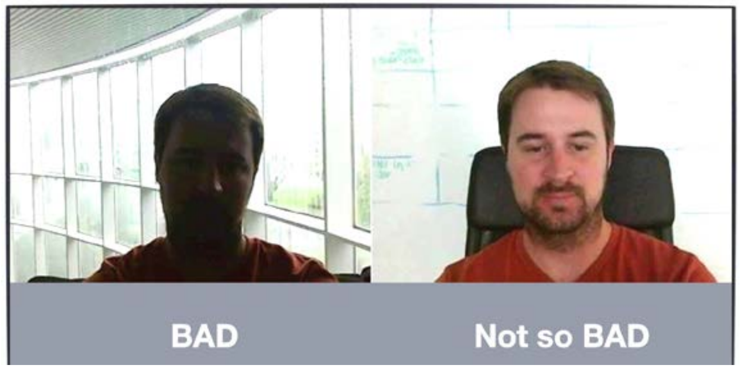
(Plays your favorite song and turns on the television.)

Why use Bio-sensing?



Impossible to continuously receive user **feedback**.

Impractical to use ego camera everywhere.



Impossible to always ensure good **illumination** conditions for the camera.

Cameras raise issues concering **privacy.**

Bio-sensing may provide the **solution!**

- **Non-intrusive**
- Does not depend on **external factors** such as illumination, occlusion, etc.
- Capable of highly **individualized** analysis.

Goals of such a Bio-sensing system

- Detect and monitor **affective** states.

- Infer **affective** states using a **minimal** number of and most **comfortable** sensors.

- Infer the **context** in **real-world** scenarios.

- Make **recommendations**/take action based on the information from above.

- Do all the above **continuously** throughout the day.

Goals of such a Bio-sensing system

- Detect and monitor **affective** states.

- Infer **affective** states using a **minimal** number of and most **comfortable** sensors.

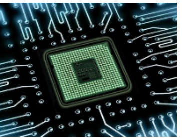
- Infer the **context** in **real-world** scenarios.

- Make **recommendations**/take action based on the information from above.

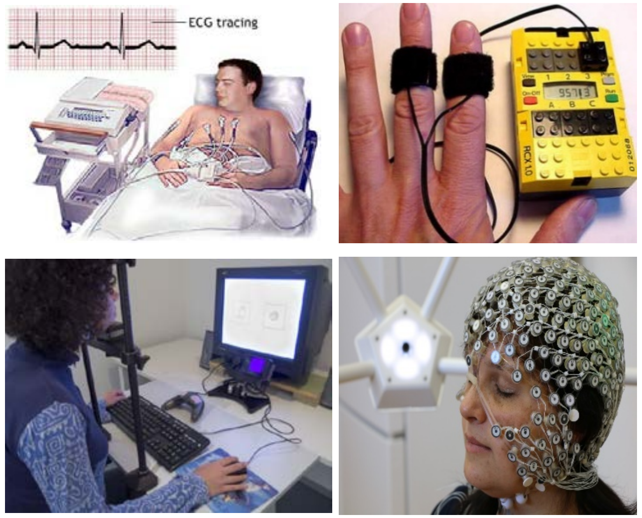
- Do all the above **continuously** throughout the day.

Five Ws and One H

- **Who** – Siddharth and collaborators
- **Where** – UC San Diego and Facebook Reality Labs
- **What** is **Affective Computing**?
- **Why** use **Bio-sensing**?
- **When** are **Multi-modal** systems advantageous?
- **How** to apply them toward **Real-world** applications?



Bio-Sensing Systems: A Brief History



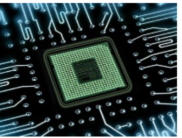
Bulky **single** modality systems[1]
(~10 years ago)

Compact **single** modality systems[2]
(~5 years ago)

Compact **multi-modal** systems[3]
(Now)

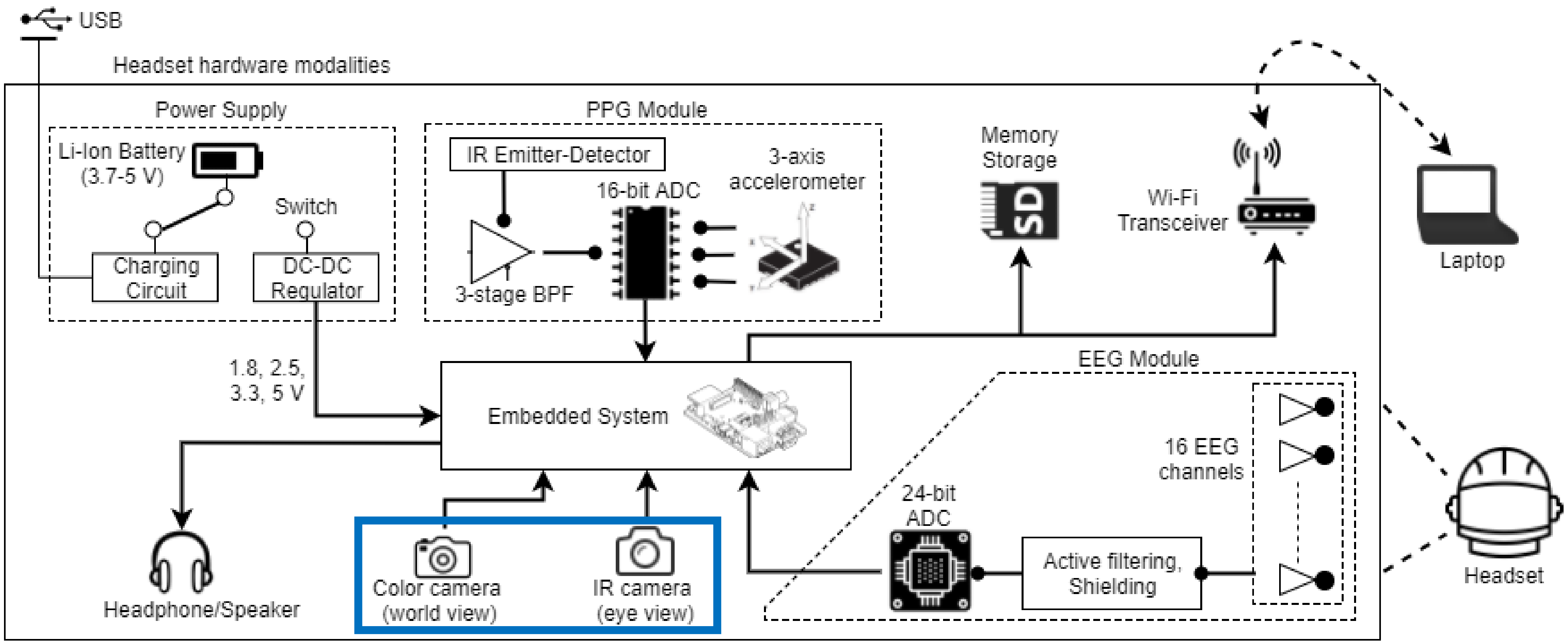
Challenges

- Do not provide **research-grade** bio-signals.
- Cannot be **customized** as per the experiment's needs.
- Data **synchronization** among sensors is cumbersome.

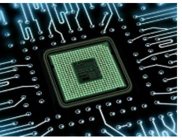


[1]https://www.sr-research.com/, https://www.brainproducts.com/
[2]https://pupil-labs.com/, https://www.emotiv.com/
[3]http://neurable.com/, http://bitalino.com/en/

OUR MULTI-MODAL BIO-SENSING SYSTEM



System Architecture

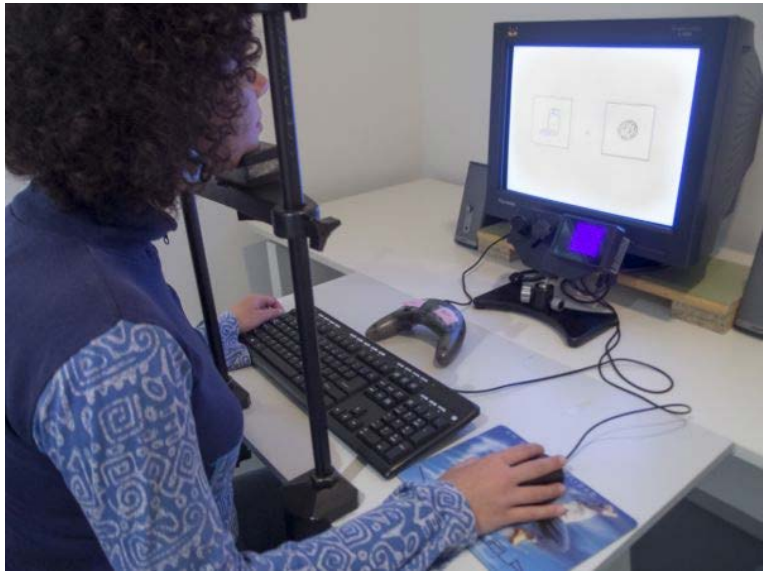


Patents filed:

Siddharth, Tzyy-Ping Jung, Terrence Sejnowski, A Wearable Multimodal Biosensing and Eye-tracking System, Provisional Patent No. 009062-8336.US00
Siddharth Siddharth, Aashish Patel, Tzyy-Ping Jung and Terrence J. Sejnowski, Wearable Multi-modal Bio-sensing System, Provisional Patent No. 62/656,890.

EYE-TRACKERS' LIMITATIONS

Tobii Eye Gaze Tracker[1]
Cost: $100

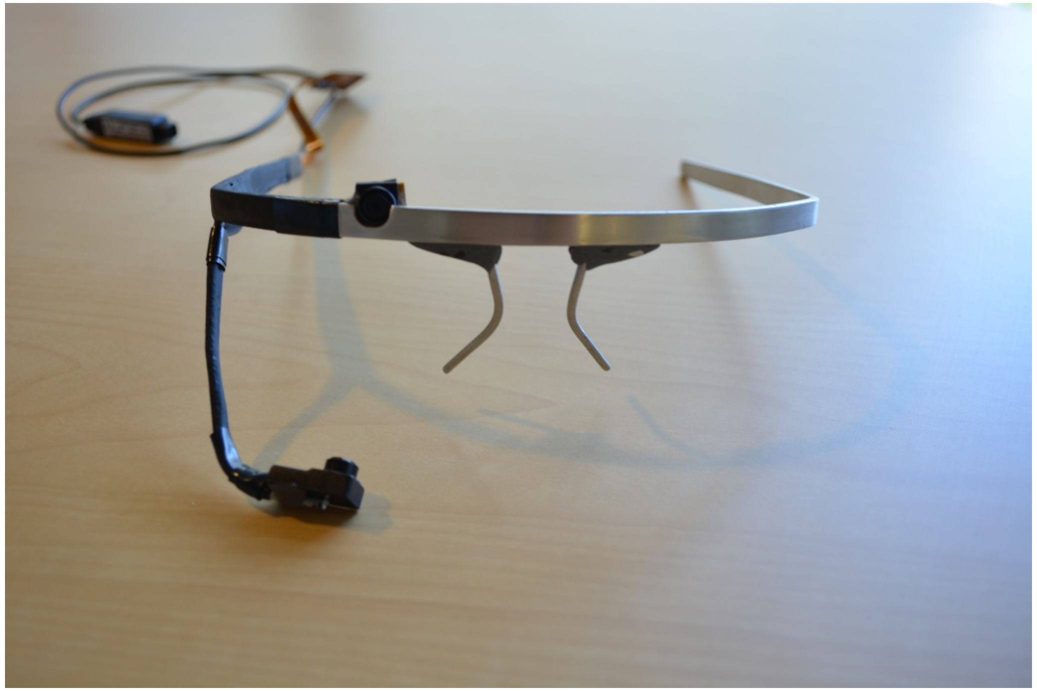


EyeLink 1000 Eye Gaze Tracker[2]
Cost: $30,000

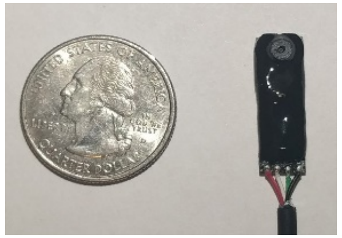
- **Non-mobile.** May even need chin rest.
- Can be very **costly.**

[1]https://tobiigaming.com/products/
[2]https://www.sr-research.com/

OUR MULTI-MODAL BIO-SENSING SYSTEM



Eye-Gaze Headset v1.0



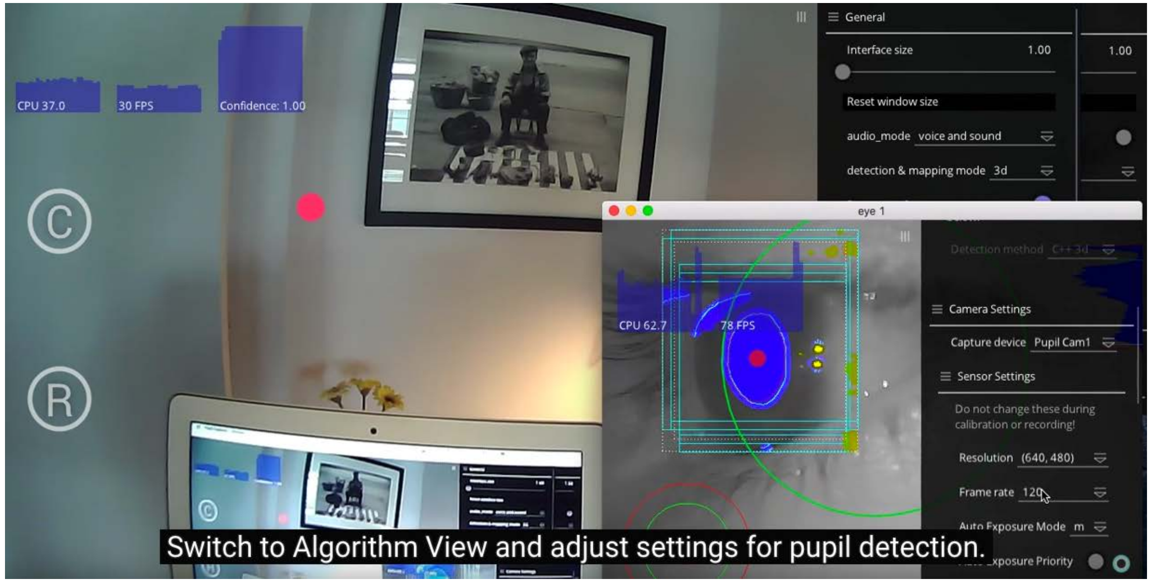
Eye Camera

Customizable Eye-Gaze Headset

- World Camera to record view from **user's perspective.**
- IR-based Eye Camera to detect **pupil.**
- Customizable headset.
- Both cameras working **simultaneously** @ 30fps and 640x480 resolution.
- Easy and **fast calibration.**[1]
- Can work while the subject is **mobile.**
- Can work in conditions with **varying illumination.**

[1]Kassner et. al., Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction, *ACM*, 2014.

OUR MULTI-MODAL BIO-SENSING SYSTEM



Eye-Gaze Software Overview

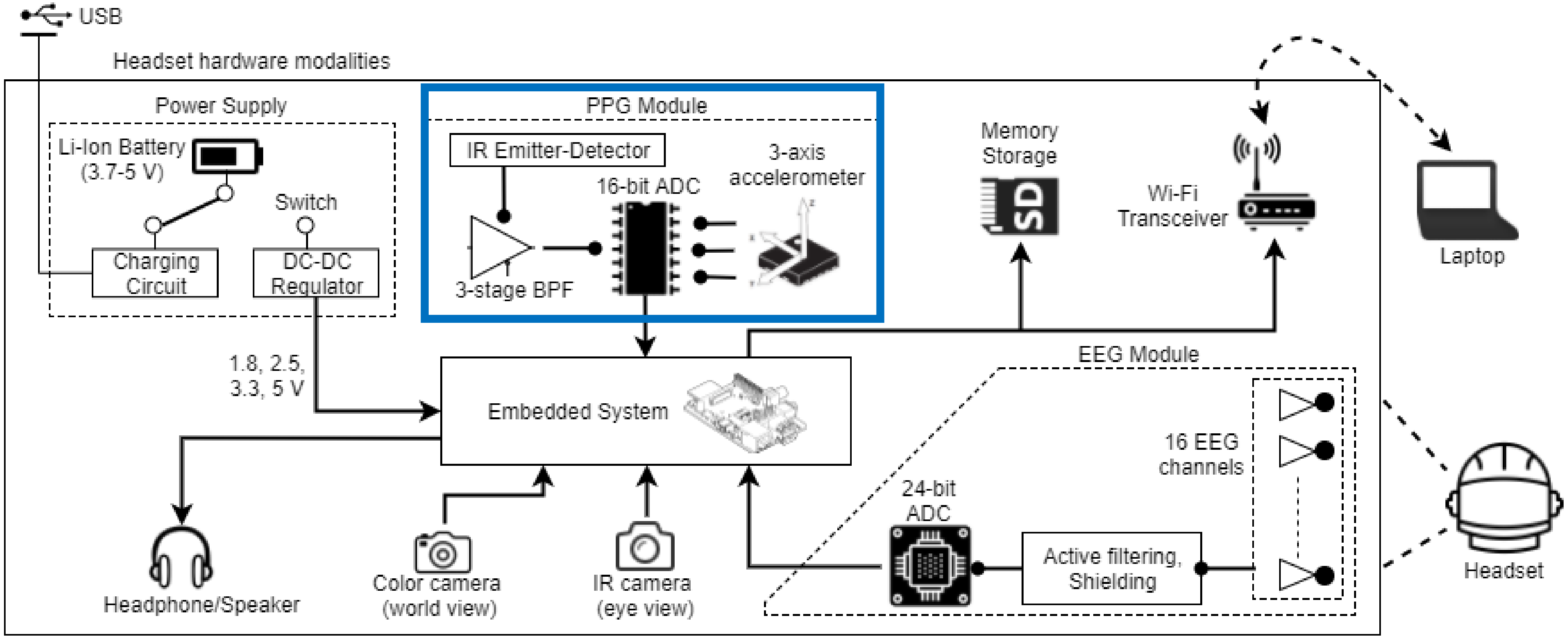
Customizable Eye-Gaze Headset

- World Camera to record view from **user's perspective.**
- IR-based Eye Camera to detect **pupil.**
- Customizable headset.
- Both cameras working **simultaneously** @ 30fps and 640x480 resolution.
- Easy and **fast calibration.**
- Can work while the subject is **mobile.**
- Can work in conditions with **varying illumination.**

Extractable Bio-Markers

- Eye-Gaze overlaid on the user's World view.
- Pupillometry (Pupil diameter, fixations, blinks, etc.)
- Pinpointing the visual stimuli to which user is affectively or sub-consciously reacting.

OUR MULTI-MODAL BIO-SENSING SYSTEM



System Architecture

Patents filed:

Siddharth, Tzyy-Ping Jung, Terrence Sejnowski, A Wearable Multimodal Biosensing and Eye-tracking System, Provisional Patent No. 009062-8336.US00

Siddharth Siddharth, Aashish Patel, Tzyy-Ping Jung and Terrence J. Sejnowski, Wearable Multi-modal Bio-sensing System, Provisional Patent No. 62/656,890.

WEARABLE CARDIAC SYSTEMS' LIMITATIONS

Zephyr BioHarness[1]

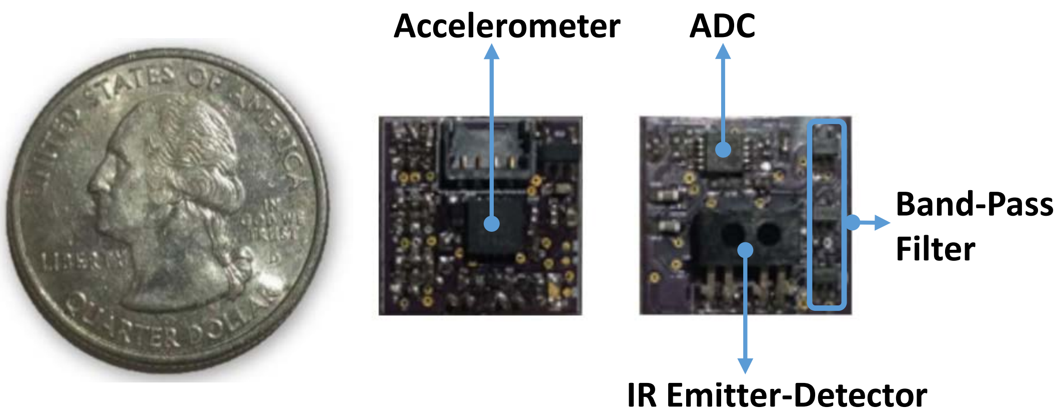
- **Difficult** and **uncomfortable** to wear.
- Require wet electrodes. So conductive gel might have to be applied.

Samsung Gear S2[2]

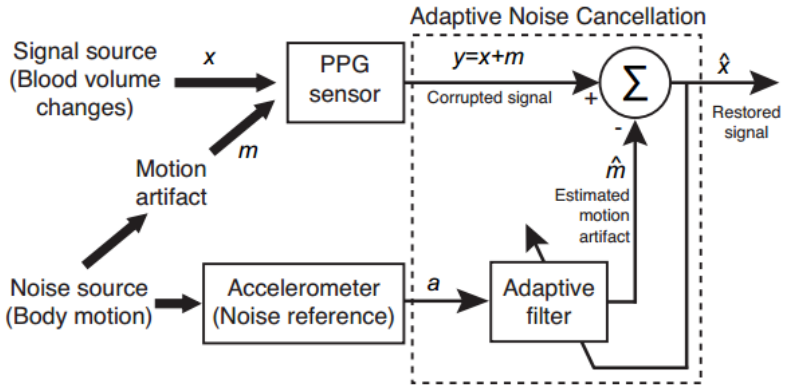
- **Low sampling rate** (usually 10Hz) to save battery power.
- Calculation of Heart-Rate Variability **(HRV)** is **not possible.**

[1]https://www.zephyranywhere.com/system/components
[2]https://www.samsung.com/global/galaxy/gear-s2/

OUR MULTI-MODAL BIO-SENSING SYSTEM



PPG Sensor Overview



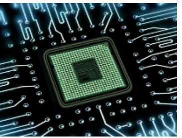
Block Diagram of ANC Configuration

Ear based Photoplethysmogram (PPG) sensor

- PPG sensor **comfortably worn** behind the ear.
- Easy to use **magnetic assembly** for physical attachment.
- IR-based (980 nm wavelength) **reflective** emitter-detector assembly.
- Three stage **band-pass** filter (0.8-4 Hz) on the board.
- Three axis **accelerometer** on the board.
- Accelerometer used to **remove noise** from PPG when the user is mobile by employing an Adaptive Noise Cancellation **(ANC)** Filter[1].
- **100 Hz.** sampling rate with **16-bit** data resolution[2].

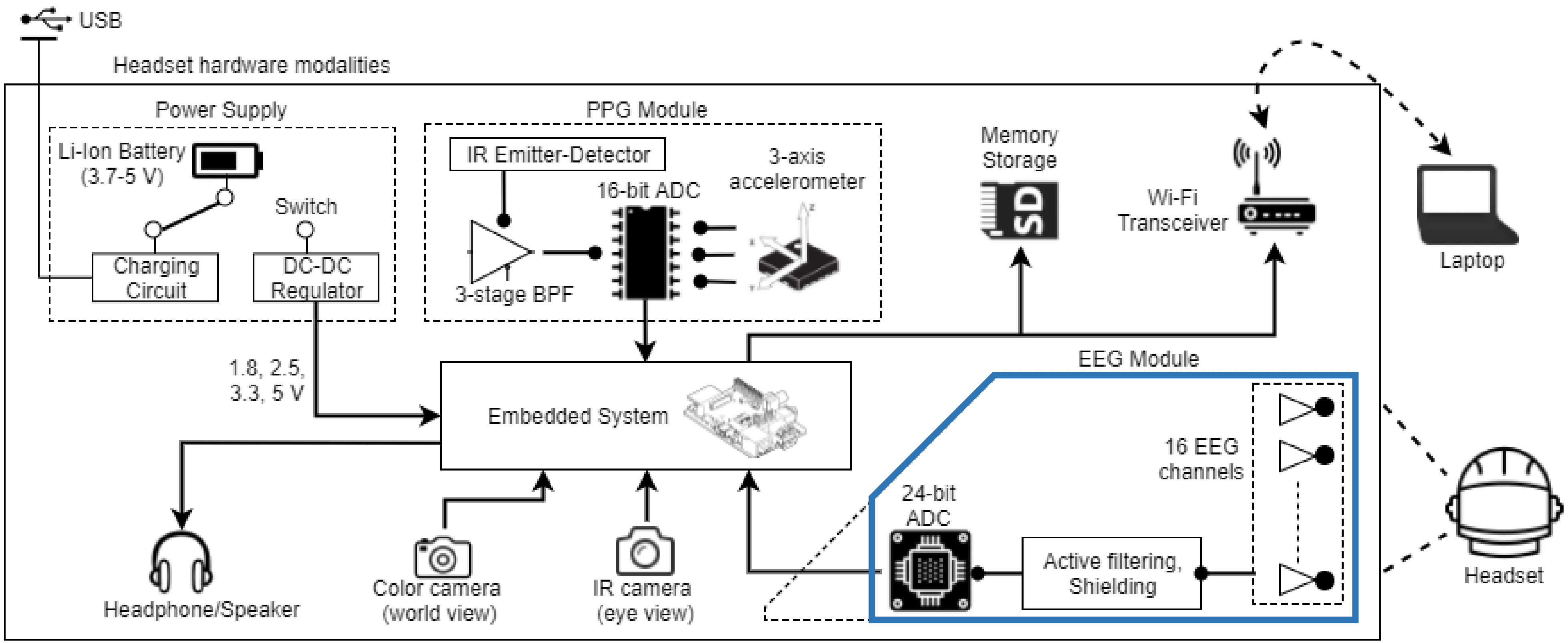
Extractable Bio-Markers

- Heart Rate
- Heart Rate Variability
- Head movement and orientation



[1]Widrow et. al., Adaptive noise cancelling: Principles and applications, *Proceedings of the IEEE*, 1975.
[2]http://www.ti.com/product/ADS1115

OUR MULTI-MODAL BIO-SENSING SYSTEM



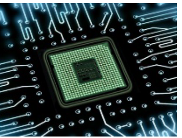
System Architecture

Patents filed:

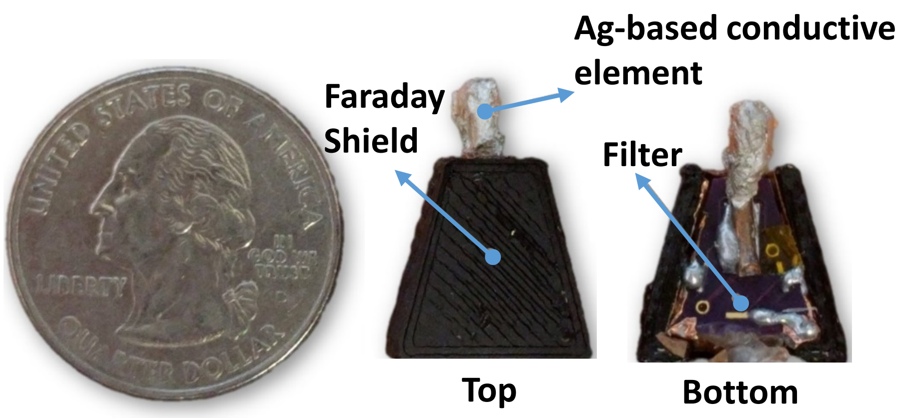
Siddharth, Tzyy-Ping Jung, Terrence Sejnowski, A Wearable Multimodal Biosensing and Eye-tracking System, Provisional Patent No. 009062-8336.US00
Siddharth Siddharth, Aashish Patel, Tzyy-Ping Jung and Terrence J. Sejnowski, Wearable Multi-modal Bio-sensing System, Provisional Patent No. 62/656,890.

Limitations of Brain-computer Interfaces (BCIs)

- Reliable BCIs are **bulky.**
- Generally use **wet electrodes.**
- Mostly **non-mobile.**
- EEG has **low spatial resolution.**
- Very **noisy.**



OUR MULTI-MODAL BIO-SENSING SYSTEM



EEG Electrode Overview

EEG Modular Unit (EMU)

- **Novel modular mechanical assembly** to penetrate hairs on the scalp.
- **Highly conductive** and low impedance electrodes made from Silver (Ag) based epoxy.
- Currently using 16 electrodes (expandable to 64).
- Completely **mobile** BCI.
- **Ultra-low noise** 24-bit ADCs being used with sampling rate up to 16 KSPS (256 SPS being used over a wireless network)[1].
- **Low-cost** ($2).
- Use of conductive shielding generates a **Faraday cage** around the sensor to shield from electromagnetic noise.

Extractable Bio-Markers

- EEG brain activity.
- Multiple secondary applications: Arousal, motor activity, visual evoked potential, speech analysis, etc.

[1]https://www.ti.com/product/ADS1299

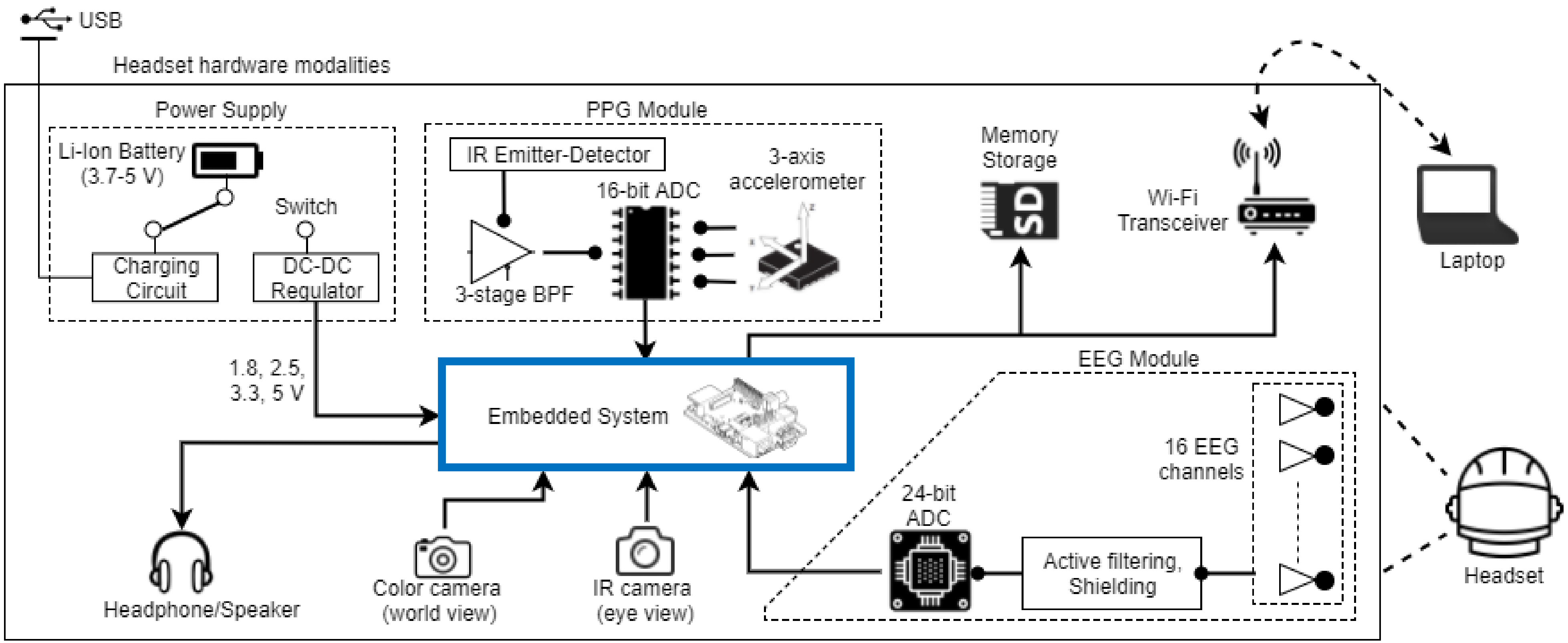
OUR MULTI-MODAL BIO-SENSING SYSTEM

Other commercially available systems that **can be integrated** as per need of the experiment:

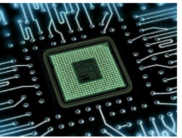
- Notch Motion-tracking System[1]
  - 3-axis IMUs on designated limbs to **track motion.**
- Microsoft Band[2]
  - Records Galvanic Skin Response **(GSR)**
- Biovotion Arm Band[3]
  - **Skin temperature** and Blood Perfusion.

[1]https://wearnotch.com/
[2]https://www.microsoft.com/en-us/band
[3]https://www.biovotion.com/

OUR MULTI-MODAL BIO-SENSING SYSTEM



System Architecture

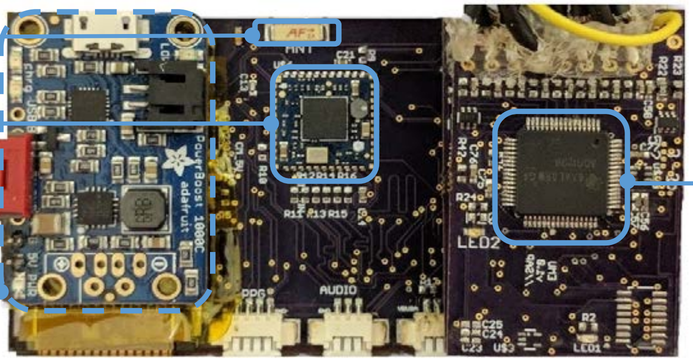


Patents filed:

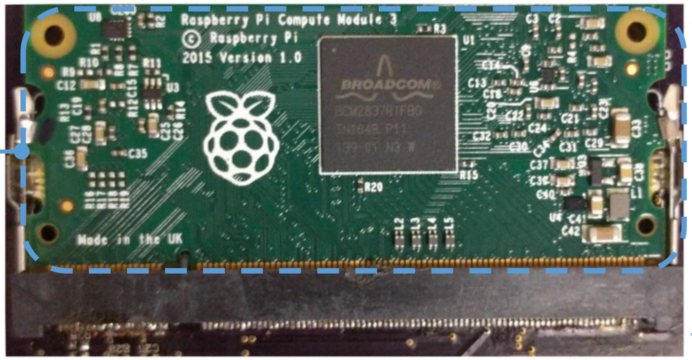
Siddharth, Tzyy-Ping Jung, Terrence Sejnowski, A Wearable Multimodal Biosensing and Eye-tracking System, Provisional Patent No. 009062-8336.US00

Siddharth Siddharth, Aashish Patel, Tzyy-Ping Jung and Terrence J. Sejnowski, Wearable Multi-modal Bio-sensing System, Provisional Patent No. 62/656,890.

OUR MULTI-MODAL BIO-SENSING SYSTEM

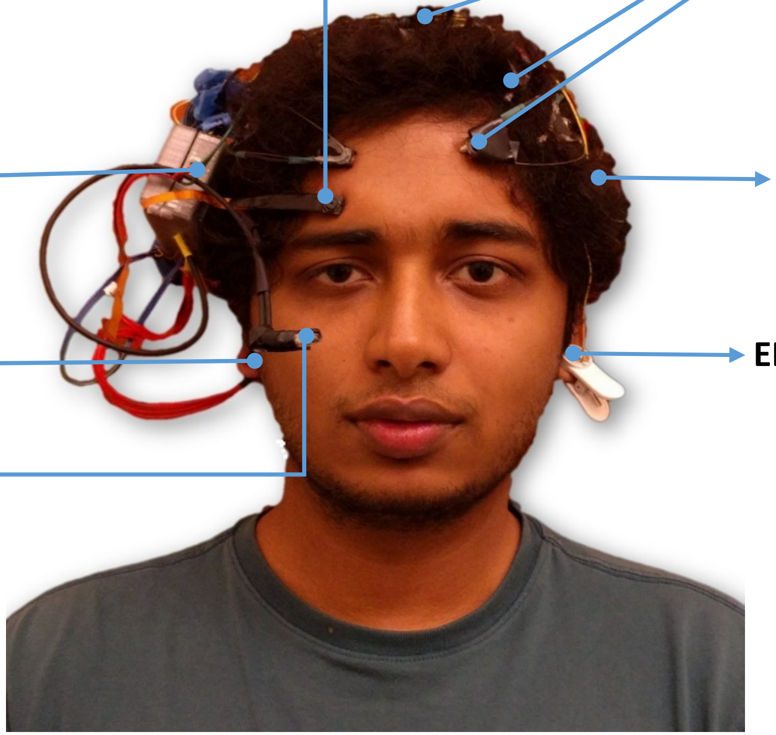


Top



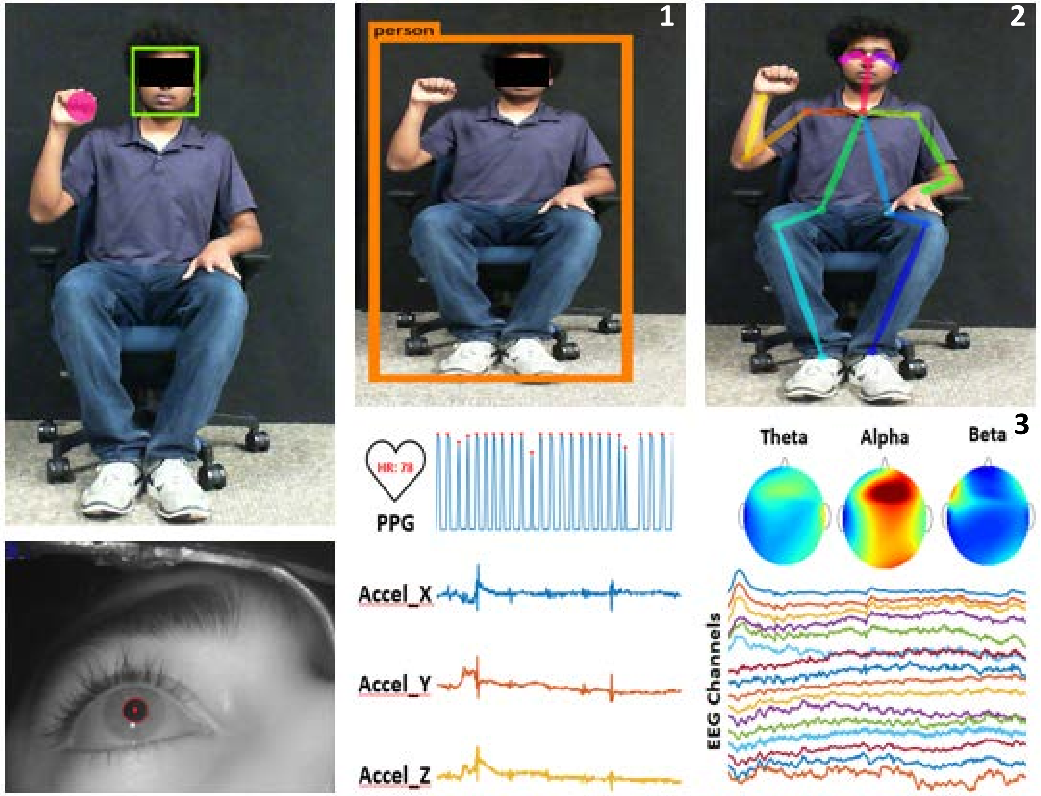
Bottom

Embedded System



Wearable Headset

CONTRIBUTIONS



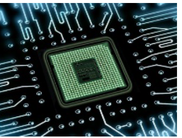
- Developed a **novel miniature** (1.6 x 1.6 cm) earlobe PPG sensor capable of signal acquisition, filtering, motion noise cancelation, **high sampling rate** (100 Hz.) and **high resolution** (16-bit) analog to digital conversion all on-board.

- Developed a **novel miniature** EEG sensor with **silver-based** Conductive element and **Faraday cage-based** shielding costing **only $2.**

- Developed a **novel eye-tracking** headset capable of measuring eye-gaze **overlaid** on the user's world view, **pupillometry**, and with the capability to work **wirelessly** rather than currently available non-mobile eye-trackers.

- Developed a **novel** miniature embedded system framework to **synchronize** and **collect** data from each of the above (and more) sensors.

[1]Redmon et. al. You only look once: Unified, real-time object detection, *IEEE CVPR,* 2016.
[2]Wei et. al., Convolutional pose machines, *IEEE CVPR,* 2016.
[3]Jung et. al.,. Removing electroencephalographic artifacts by blind source separation, *Psychophysiology*, 2000.



SYSTEM EVALUATION

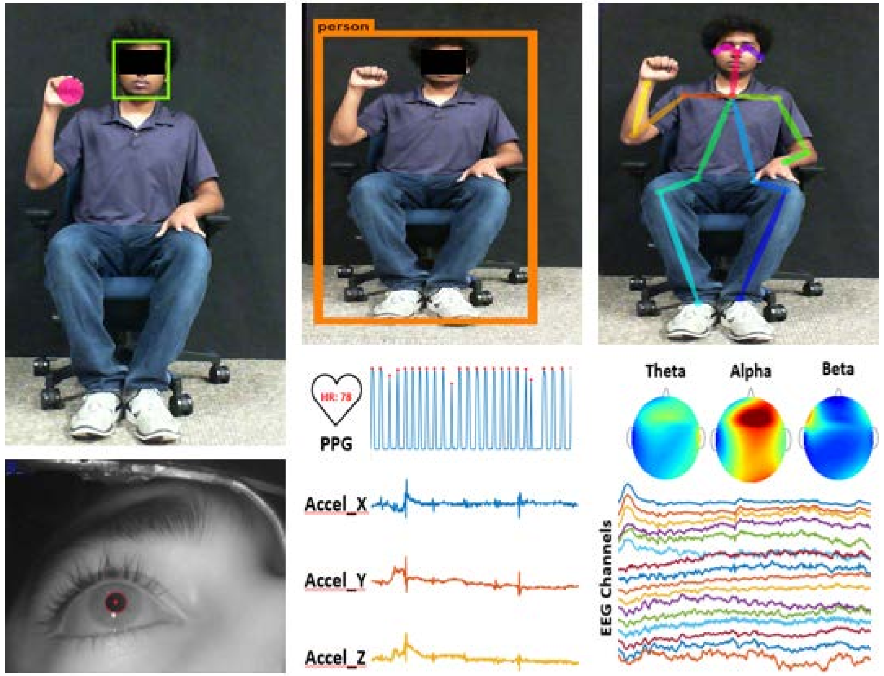
Where is Waldo?

- Allows for studying **EEG** with true and false **gaze fixations.**

- 10 subjects
- 13 Waldo scenes
- 50 RPS trials.

- **Real-world** tasks but somewhat **“controlled”.**

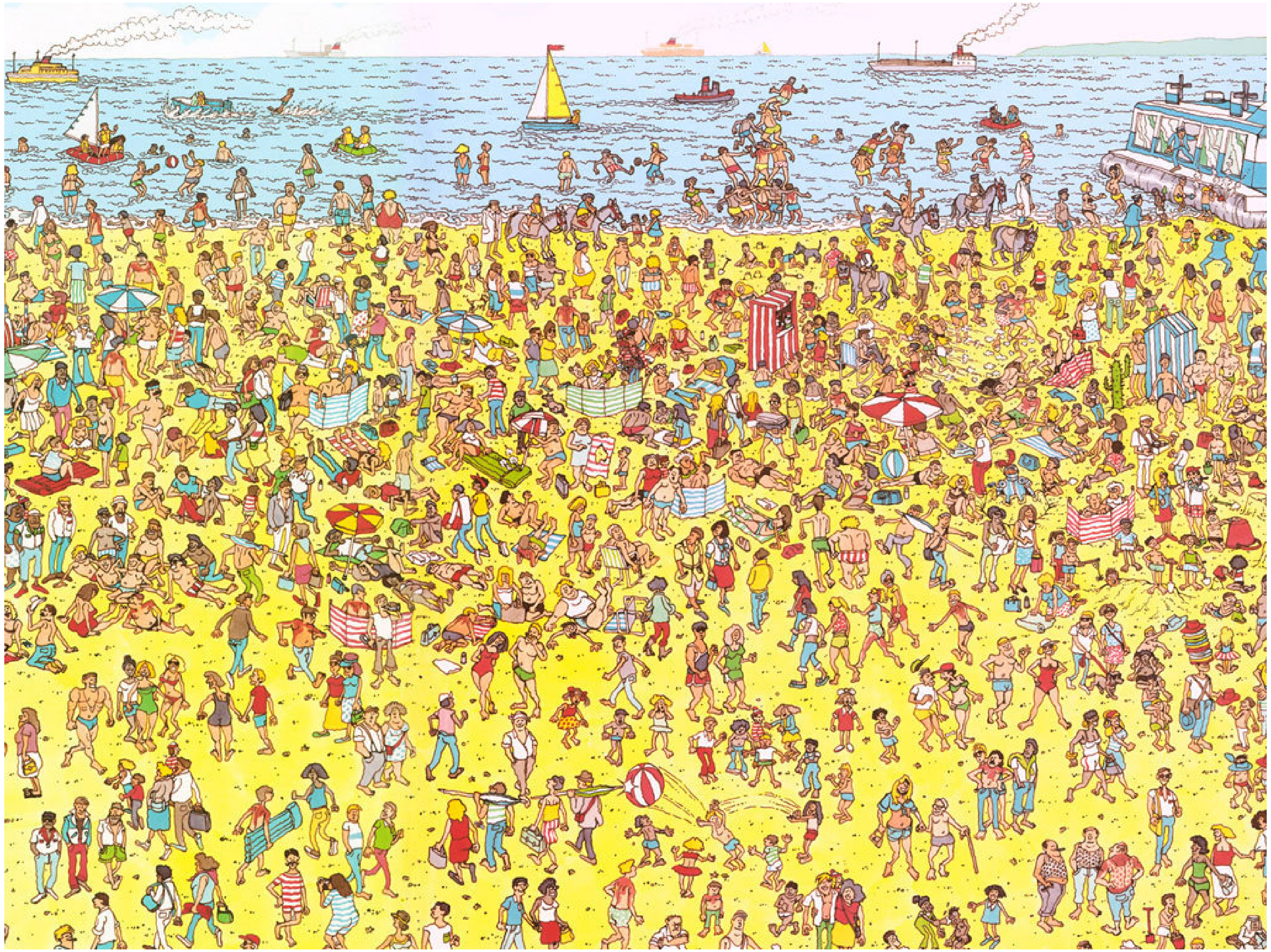
Rock-Paper-Scissors (RPS)



- Allows for studying **win/loss** type of mood without subject’s **direct feedback** after each trial.

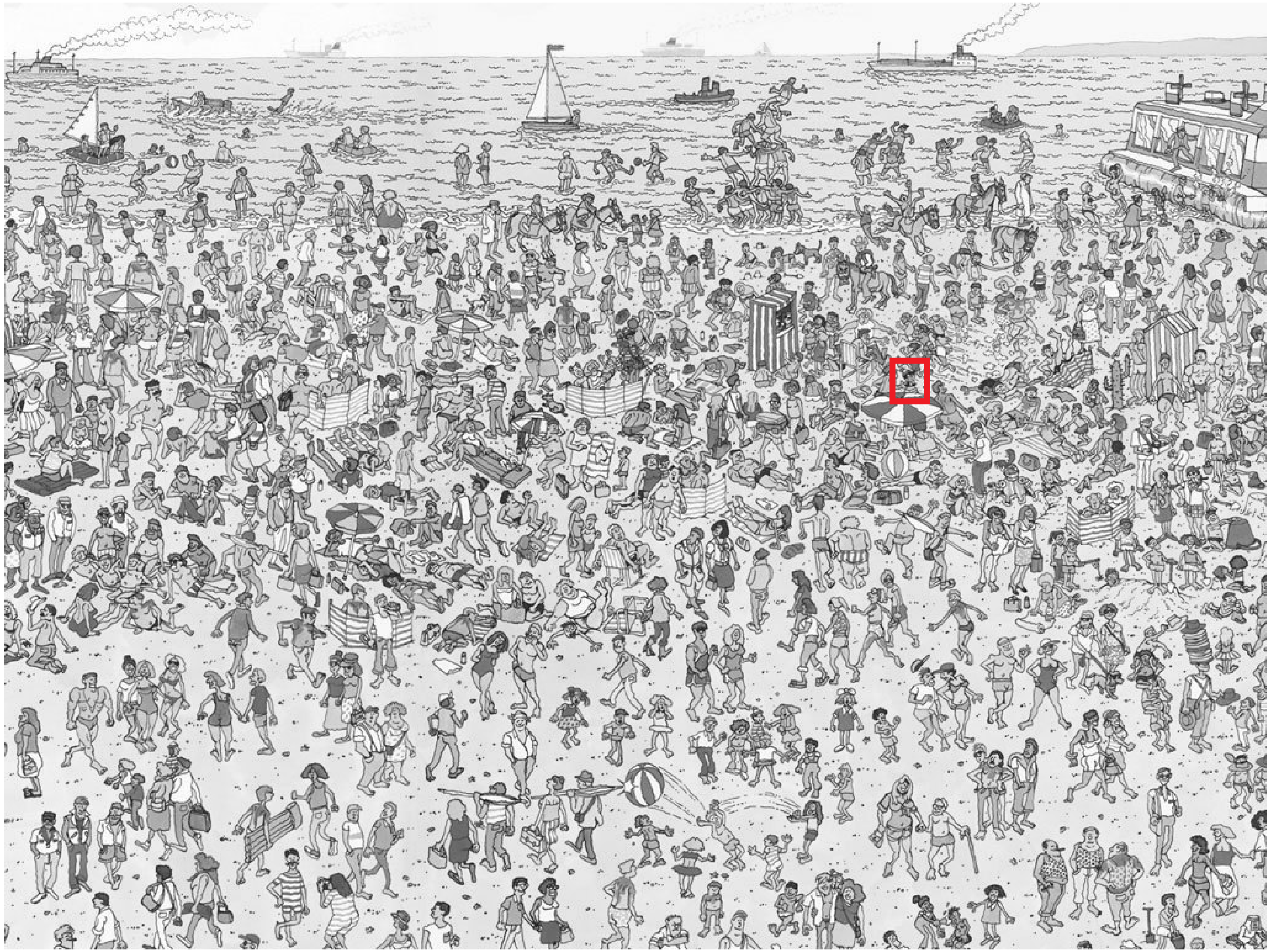
MULTI-MODAL EVALUATION

Where is Waldo?



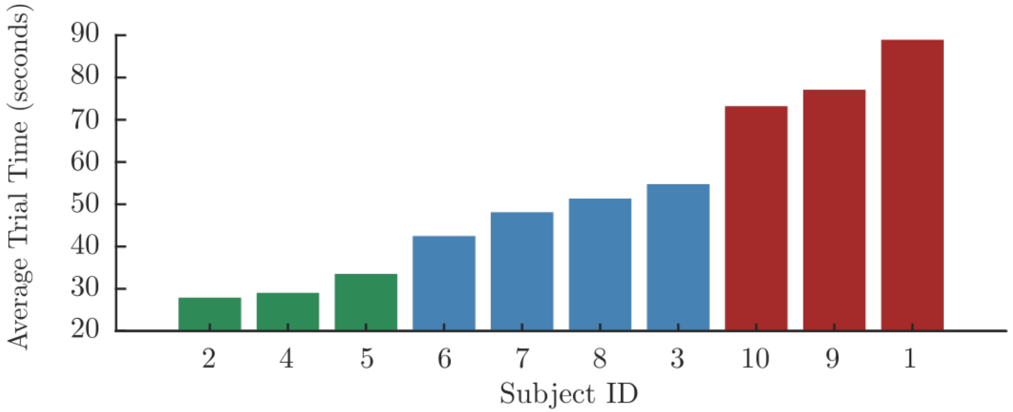
MULTI-MODAL EVALUATION

Where is Waldo?



MULTI-MODAL EVALUATION

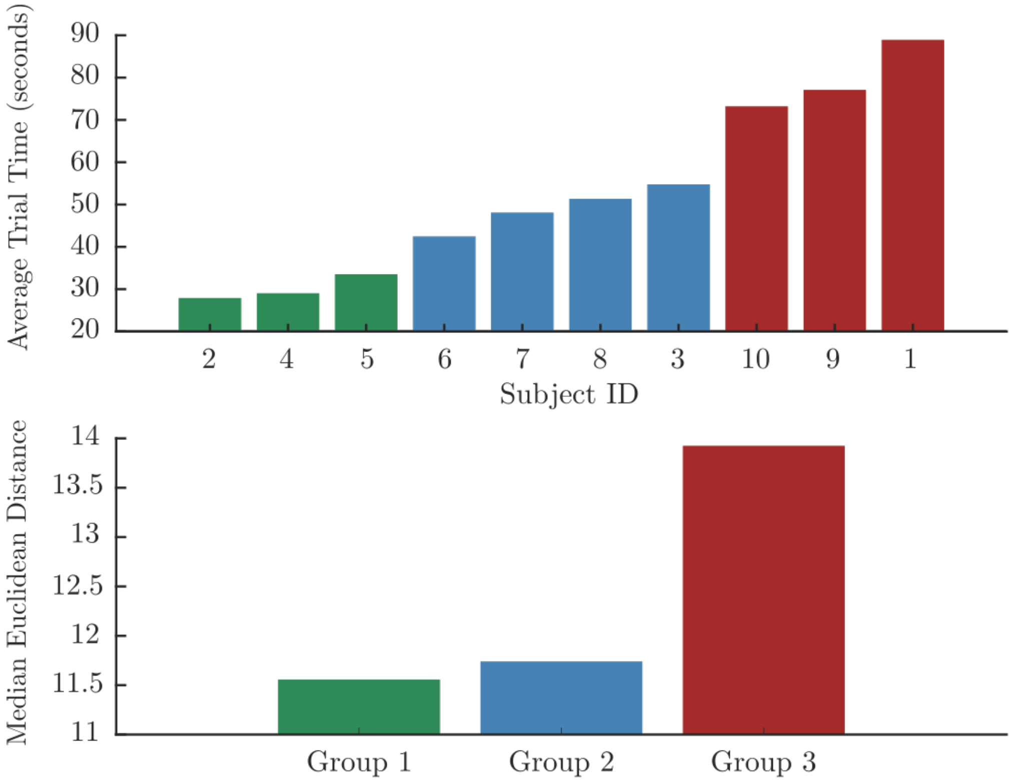
Where is Waldo?



- Forming three clusters based on how much **time on average** subjects take to complete the Waldo experiment.

MULTI-MODAL EVALUATION

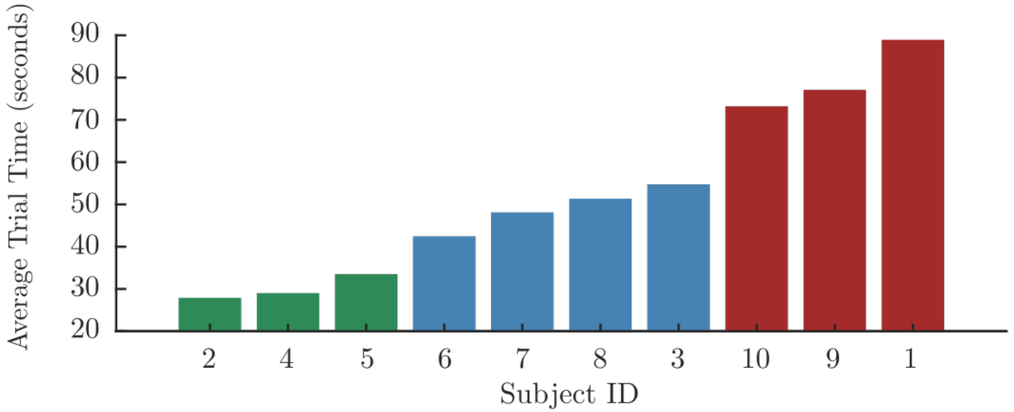
Where is Waldo?

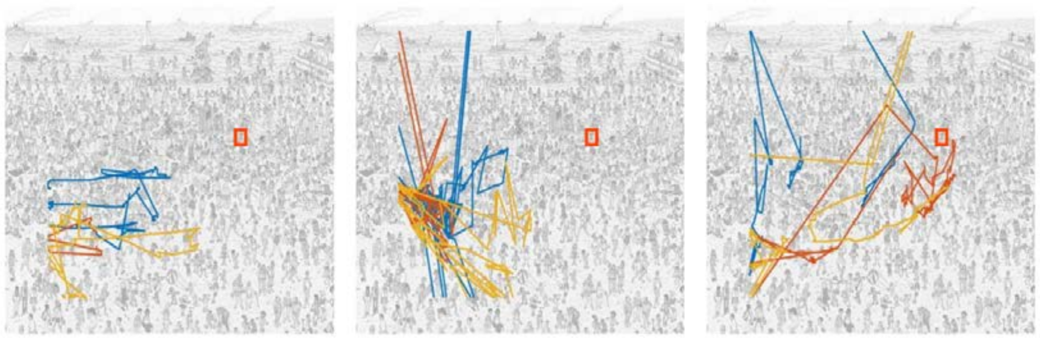


- Forming three clusters based on how much **time on average** subjects take to complete the Waldo experiment.
- Finding the **median Euclidean distance** between successive fixations across all fixations by the subjects in that cluster.
- Fixation was defined as to be minimum 500ms long and 25 pixels as the **maximum inter-sample** Euclidean distance.

MULTI-MODAL EVALUATION

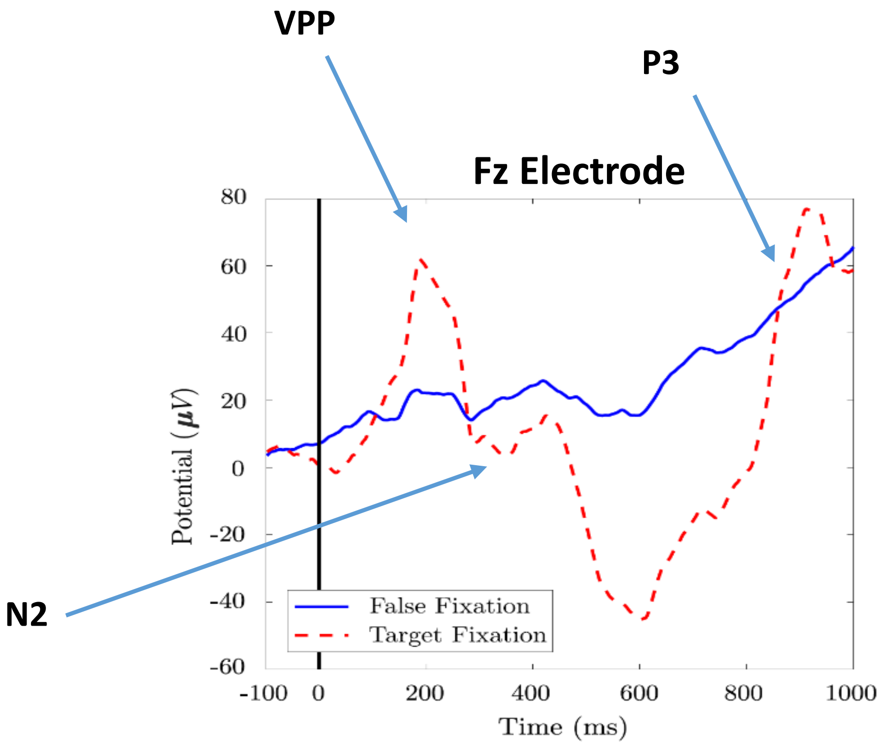
Where is Waldo?





- Forming three clusters based on how much **time on average** subjects take to complete the Waldo experiment.
- Finding the **median Euclidean distance** between successive fixations across all fixations by the subjects in that cluster.
- Fixation was defined as to be minimum 500ms long and 25 pixels as the **maximum inter-sample** Euclidean distance.
- Subjects who tend to search for Waldo randomly across the page tend to **take longer** than the subjects who search in small portions of the visual area.

MULTI-MODAL EVALUATION



EEG Fixation-related Potential

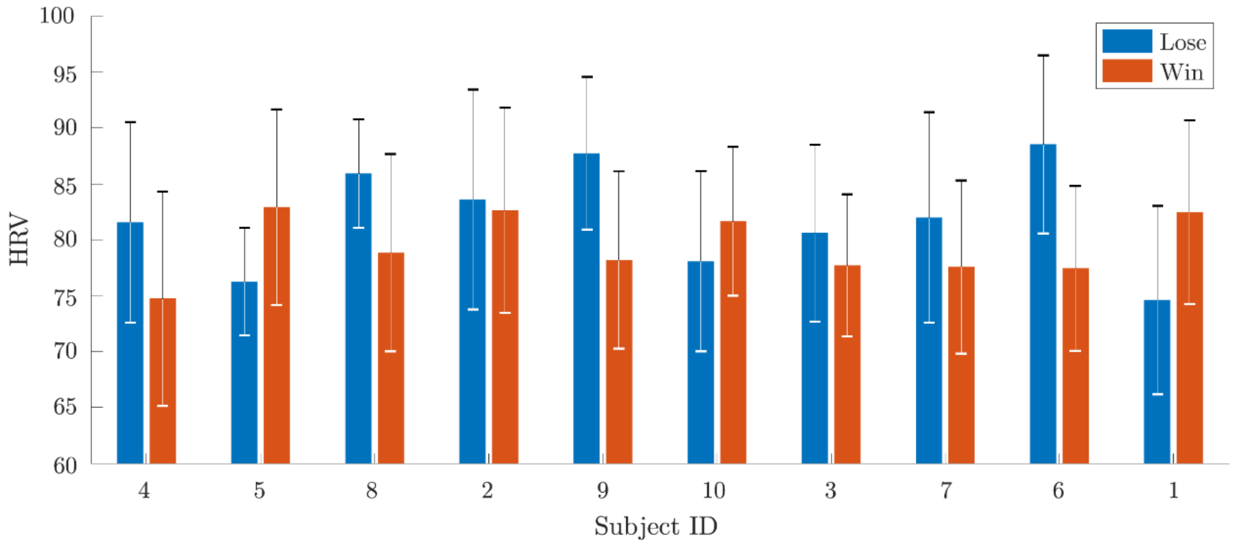
Where is Waldo?

- Large peak at 200ms i.e. VPP and the occurrence of N2 are **consistent with earlier findings** that VPP and N2 are associated with face stimuli (Wang et al.[1], Kaufmann et al.[2]).
- Large P3 associated with **decision-making** is clearly much larger for targets than non-targets (Polich et al.[3]).
- The slightly smeared nature of the P3 response is likely due to the fact that the latency of the P3 can **vary across trials** and individuals and the fixation-related potentials (FRPs) are time-locked to the onset of fixation.

[1] Wang et. al., Convolutional Neural Network for Target Face Detection using Single-trial EEG Signal, *IEEE EMBC*, 2018.
[2] Kauffman et. al., N250 ERP correlates of the acquisition of face representations across different images, *Journal of Cognitive Neuroscience*, 2009.
[3] Polich et. al., Updating P300: an integrative theory of P3a and P3b, *Clinical neurophysiology*, 2007.

MULTI-MODAL EVALUATION

Rock-Paper-Scissors



- Computing HRV using **pNN50**[1] measure across all trials.
- Clearly HRV **shows correlation** between losing and winning trials across all subjects.

[1]Hutchinson et. al., Statistics and graphs for heart-rate variability: pNN50 or pNN20, *Physiology Measurement*, 2003.

MULTI-MODAL EVALUATION

Rock-Paper-Scissors

MODALITY PERFORMANCE FOR MULTI-MODAL CLASSIFICATION

Subject ID	1	2	3	4	5	6	7	8	9	10	Mean	Max	Std.
Classification Performance (Loss/Draw/Win) Chance Accuracy: 33%													
EEG (1-sec)	56	56	52	54	62	56	54	46	52	50	**53.80**	**62**	**4.26**
PPG (15-sec)	58	58	60	46	46	48	54	58	56	52	**53.60**	**60**	**5.32**
EEG + PPG (15-sec)	54	54	52	52	56	54	56	52	54	54	**53.80**	**56**	**1.48**
Classification Performance (Loss/Win) Chance Accuracy: 50%													
EEG (1-sec)	87.88	80.65	86.84	70.97	63.33	81.82	72.73	70.00	68.97	72.41	**75.56**	**87.88**	**8.21**
PPG (15-sec)	87.88	87.10	86.84	70.97	70.00	81.82	75.76	86.67	75.86	72.41	**79.53**	**87.88**	**7.30**
EEG + PPG (15-sec)	84.85	87.10	81.58	80.65	70.00	81.82	72.73	73.33	68.97	68.97	**77.00**	**87.10**	**6.92**

Leave one subject out validation was performed. All values denote percentage accuracy.

- **Leave-one-subject-out** cross validation.
- **Conditional Entropy** features used for EEG.
- HRV and Statistical features used for PPG.
- Extreme Learning Machines (ELM) used for **classification.**
- Both modalities tend to work well at different **temporal resolutions.**
- **Combining** the modalities decreases the standard deviation across the subjects.

CONTRIBUTIONS

- Evaluated the designed sensor platform on **practical "real-world" tasks** to demonstrate the **advantage** of simultaneously using a **multi-modal bio-sensing** system. To this end, a framework was designed to **learn information** from individual sensor modalities and use their **fusion** for evaluating performance.

- It was **impossible** to garner such **fundamental insights** into the strategies employed by users during such **"real-world"** tasks without a **multi-modal bio-sensing** system. Thus, such systems should be used when a single modality cannot capture the underlying **physiology.**

Five Ws and One H

- **Who** – Siddharth and collaborators
- **Where** – UC San Diego and Facebook Reality Labs
- **What** is **Affective Computing**?
- **Why** use **Bio-sensing**?
- **When** are **Multi-modal** systems advantageous?
- **How** to apply them toward **Real-world** applications?

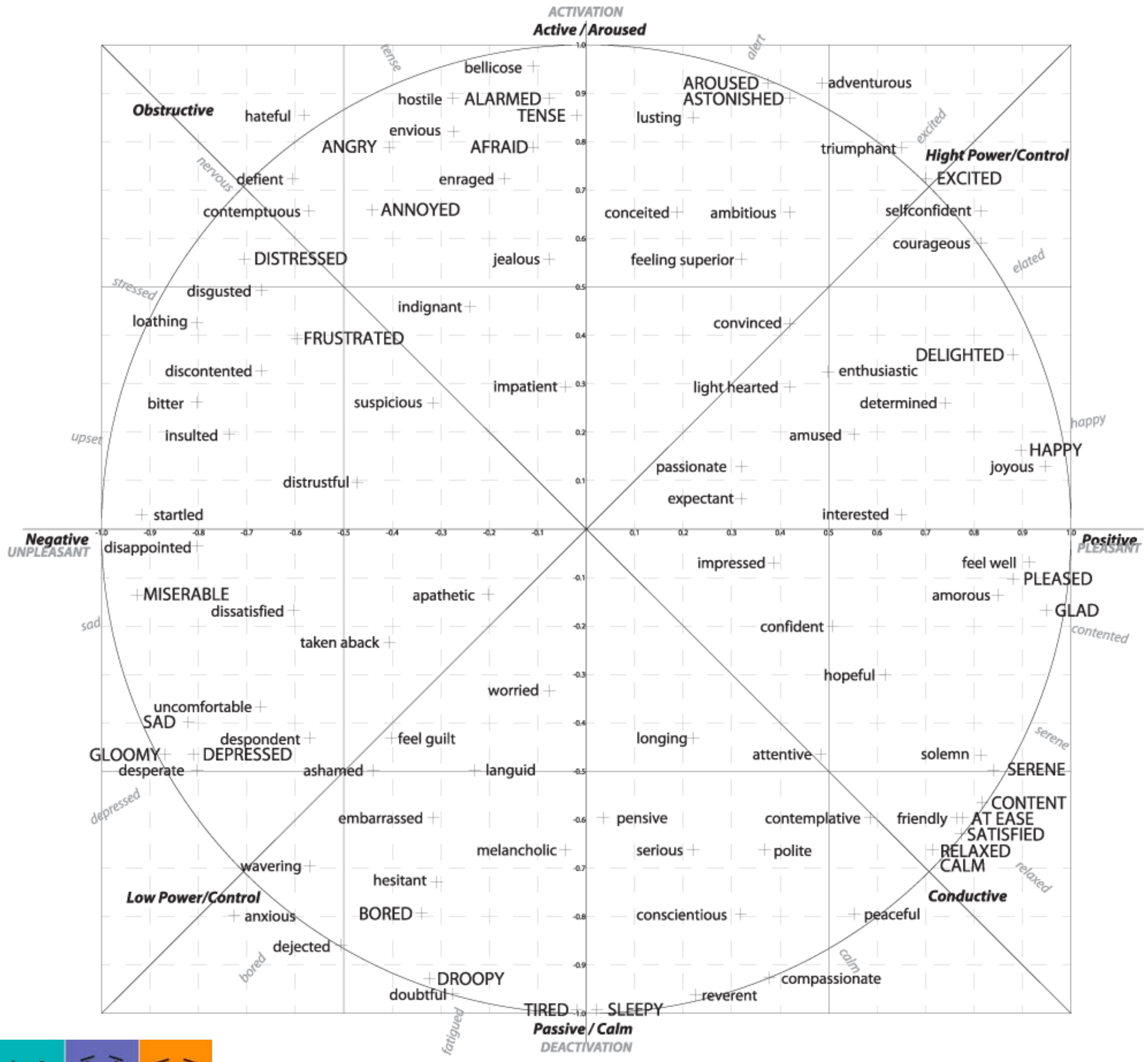
How to apply them toward **Real-world** applications?

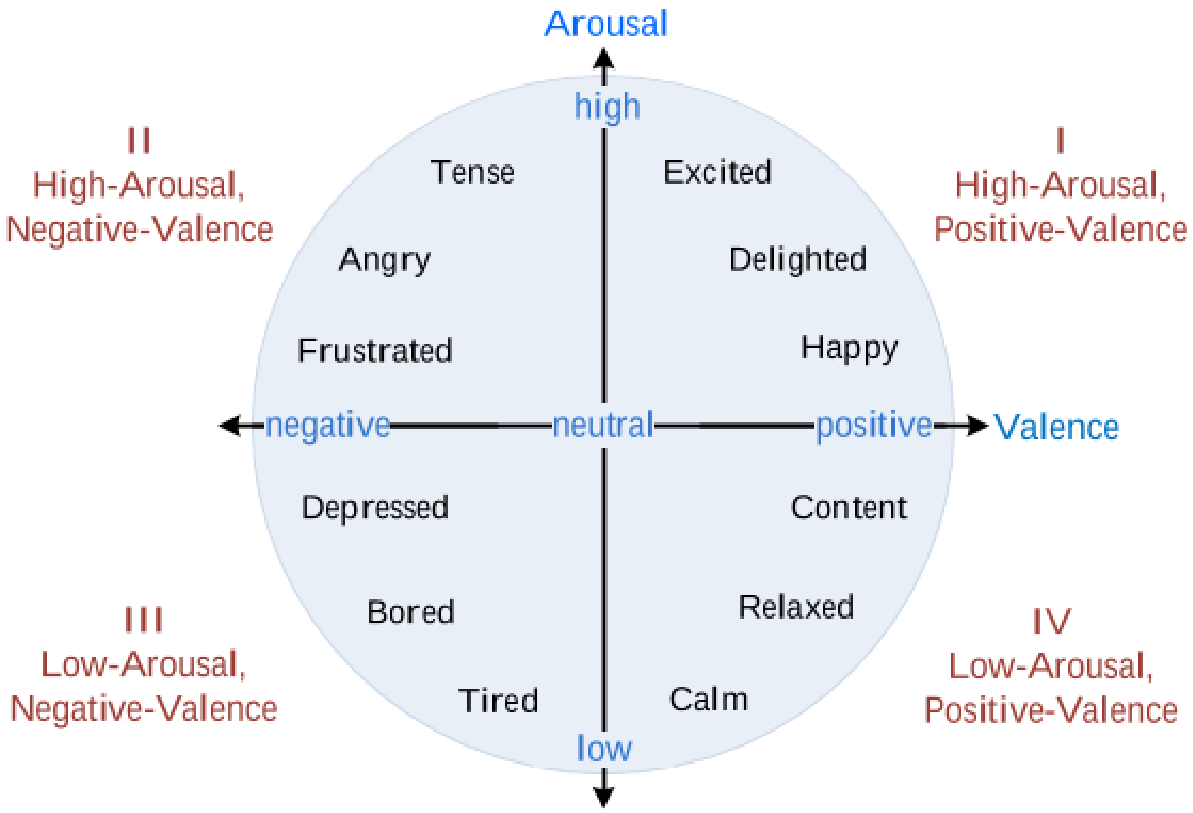
- Consuming Multimedia Content
- Monitoring Driver Awareness

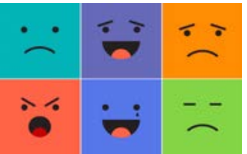
How to apply them toward **Real-world** applications?

- Consuming Multimedia Content
- Monitoring Driver Awareness

EMOTION CIRCUMPLEX MODEL







Russell, J.A., A circumplex model of affect, *Journal of personality and social psychology*, 39(6), p. 1161, 1980.

CONSUMING MULTIMEDIA CONTENT

MAHNOB-HCI Dataset[1]

- 27 subjects
- 20 **short** (0.5-2.5 minutes long) movie clips.

- Data includes:

a) Upper Body 2D **videos**
b) 32 channel Electroencephalogram **(EEG)**
c) 1 channel Electrocardiogram **(ECG)**
d) 1 channel Galvanic Skin Response **(GSR)**
e) Eye-gaze

- **User-reported** affective states:

a) Valence (ranging from 1 to 9)
b) Arousal (ranging from 1 to 9)
c) Emotion (divided into 12 classes)
d) Happiness….

[1]Soleymani et. al., A multimodal database for affect recognition and implicit tagging, *IEEE Transactions on Affective Computing*, 2012.

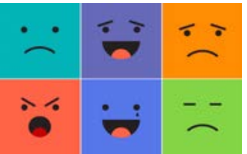
EXAMPLE MULTIMEDIA CLIP

EXAMPLE MULTIMEDIA CLIP

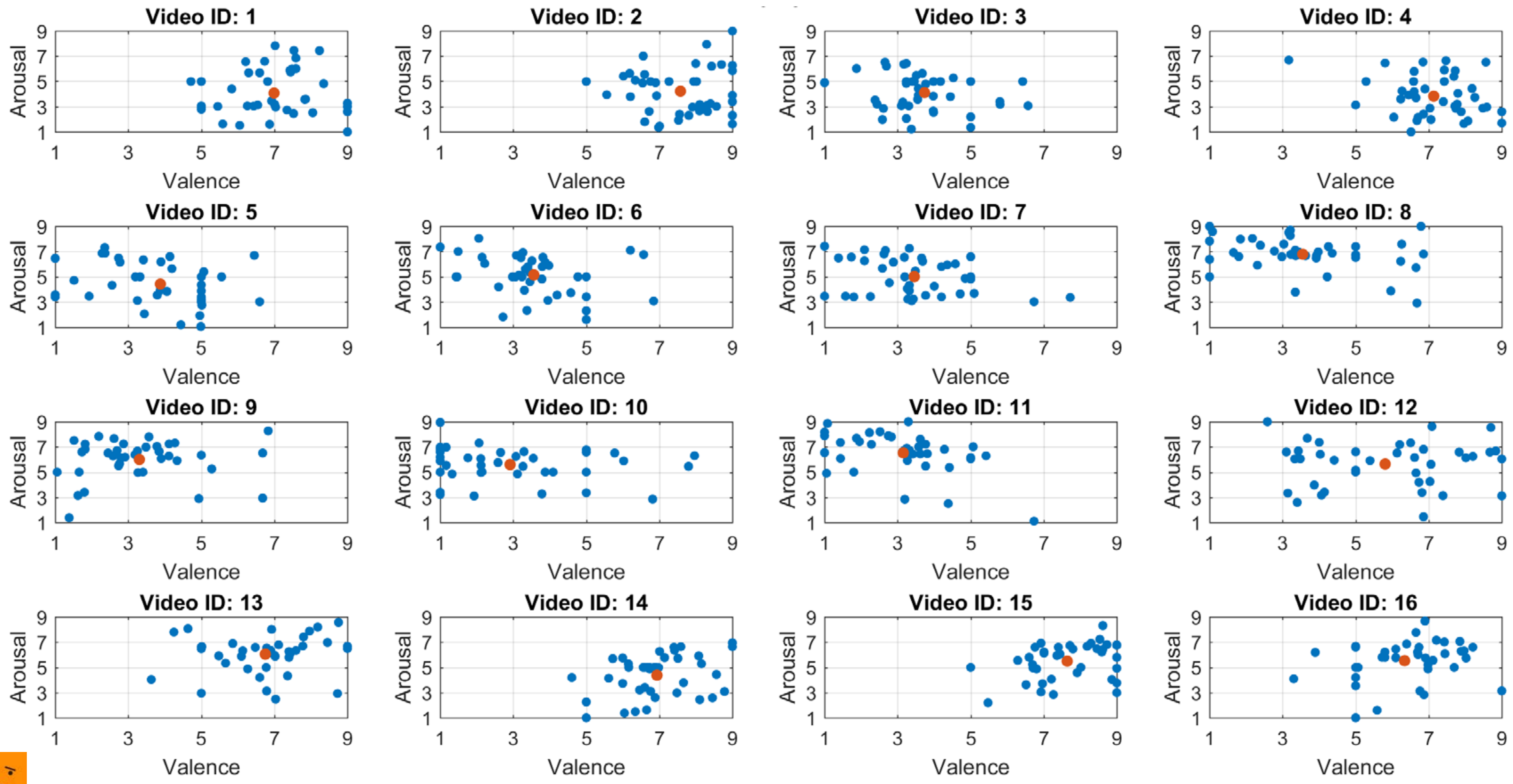
PREVIOUS WORK

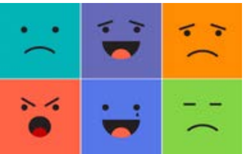
Study	Used Modalities	Extracted Features	Classifier	Evaluation
			DEAP Dataset	
Liu et al. [28]	EEG	Fractal dimension (FD) based	SVM	Only 22 of the 32 subjects used. 50.8% Valence (4-classes) and 76.51% Arousal/Dominance.
Yin et al. [34]	EEG, ECG, EOG, GSR, EMG, Skin temperature, Blood volume, Respiration	Various	MESAE	77.19% Arousal and 76.17% Valence (2-classes) using fusion of all modalities.
Patras et al. [30]	EEG	PSD	Bayesian Classifier	62% Valence and 57.6% Arousal (2-classes)
Chung et al. [36]	EEG	Various	Bayesian weighted-log-posterior	70.9% Valence and 70.1% Arousal (2-classes)
Shang et al. [37]	EEG, EOG, EMG	Raw data	Deep Belief Network, Bayesian Classifier	51.2% Valence, 60.9% Arousal, and 68.4% Liking (2-classes)
Campos et al. [38]	EEG	Various	Genetic algorithms, SVM	73.14% Valence and 73.06% Arousal (2-classes)
			AMIGOS Dataset	
Miranda et al. [31]	EEG, ECG, GSR	Various	SVM	*57.6/53.1/53.5/57 Valence and 59.2/54.8/55/58.5 Arousal (2-classes) using EEG/GSR/ECG alone/EEG, GSR, and ECG fusion.
			MAHNOB-HCI Dataset	
Soleymani et al. [32]	EEG, ECG, GSR, Respiration, Skin Temperature	Various	SVM	57/45.5/68.8/76.1% Valence and 52.4/46.2/63.5/67.7% Arousal (2-classes) using EEG/Peripheral/Eye gaze/Fusion of EEG and gaze.
Koelstra et al. [39]	EEG, Faces	Various	Decision classifiers fusion	73% Valence and 68.5% Arousal (2-classes) using EEG and Faces fusion.
Alasaarela et al. [40]	ECG	Various	KNN	59.2% Valence and 58.7% Arousal (2-classes)
Zhu et al. [41]	EEG and Video stimulus	Various	SVM	55.72/58.16% Valence and 60.23/61.35% Arousal (2-classes) for EEG alone/Video stimulus as privileged information with EEG.
			DREAMER Dataset	
Stamos et al. [33]	EEG, ECG	PSD, HRV	SVM	62.49/61.84% Valence and 62.17/62.32% Arousal (2-classes) using EEG alone/EEG and ECG fusion.

*Denotes mean F1-score. Accuracy value not available.



BUT, EMOTIONS ARE HIGHLY INDIVIDUALISTIC

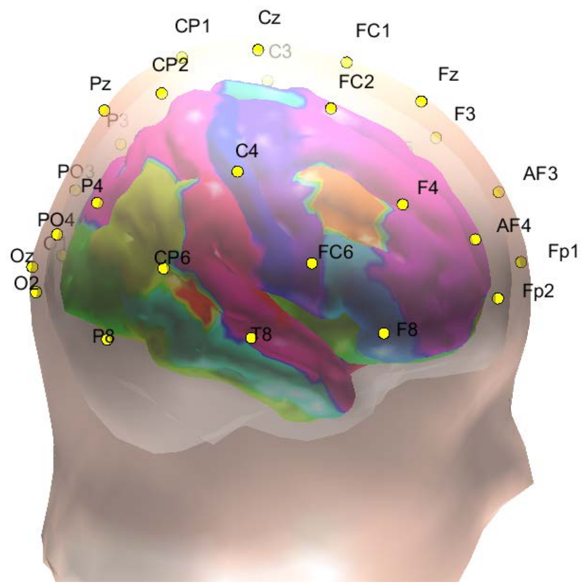


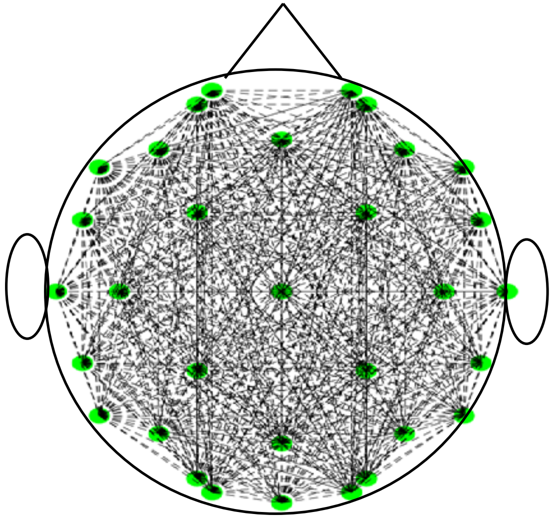


BUT, DISCREPANCIES AMONG DATASETS

DEAP Dataset	AMIGOS Dataset	MAHNOB-HCI Dataset	DREAMER Dataset
32 subjects	40 subjects	27 subjects	23 subjects
40 trials using music videos (trial length fixed at 60 seconds)	16 trials using movie clips (trial length varying between 51 and 150 seconds)	20 trials using movie clips (trial length varying between 34.9 and 117 seconds)	18 trials using movie clips (trial length varying between 67 and 394 seconds)
Raw and pre-processed data available	Raw and pre-processed data available	Only raw data available	Only raw data available
32-channel EEG system (Two different EEG systems used. Channel locations: Fp1, AF3, F7, F3, FC1, FC5, T7, C3, CP1, CP5, P7, P3, Pz, PO3, O1, Oz, O2, PO4, P4, P8, CP6, CP2, C4, T8, FC6, FC2, F4, F8, AF4, Fp2, Fz, Cz)	14-channel EEG system (A single EEG system used for all subjects. Channel locations: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4)	32-channel EEG system (A single EEG system used for all subjects. Channel locations: Fp1, AF3, F7, F3, FC1, FC5, T7, C3, CP1, CP5, P7, P3, Pz, PO3, O1, Oz, O2, PO4, P4, P8, CP6, CP2, C4, T8, FC6, FC2, F4, F8, AF4, Fp2, Fz, Cz)	14-channel EEG system (A single EEG system used for all subjects. Channel locations: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4)
—	2-channel ECG system	3-channel ECG system	2-channel ECG system
1-channel PPG system	—	—	—
1-channel GSR system	1-channel GSR system	1-channel GSR system	—
Face video recorded for 22 of 32 subjects (EEG cap and EOG electrodes occludes parts of the forehead and cheeks)	Face video recorded for all subjects (Only a small portion of the forehead is occluded by the EEG system)	Face video recorded for all subjects (Only a small portion of the forehead is occluded by the EEG system)	—
3-seconds of pre-trial baseline data available.	No baseline data available.	30 seconds of pre-trial and post-trial baseline data available.	61 seconds of pre-trial baseline data available
Valence/Arousal/Liking rated using a continuous scale between 1 to 9	Valence/Arousal/Liking rated using a continuous scale between 1 to 9	Valence/Arousal rated using a discrete scale of integers from 1 to 9	Valence/Arousal rated using a discrete scale of integers from 1 to 5
Koelstra et al., DEAP: A database for emotion analysis using physiological signals, *IEEE Transactions on Affective Computing*, 2012.	Miranda-Correa et al. AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups, *IEEE TAC*, 2017.	Soleymani et al., A multimodal database for affect recognition and implicit tagging, *IEEE Transactions on Affective Computing*, 2012.	Katsigiannis et al., DREAMER: A database for emotion recognition through EEG and ECG, *IEEE journal of biomedical and health informatics*, 2018.

MULTI-MODAL DATA ANALYSIS





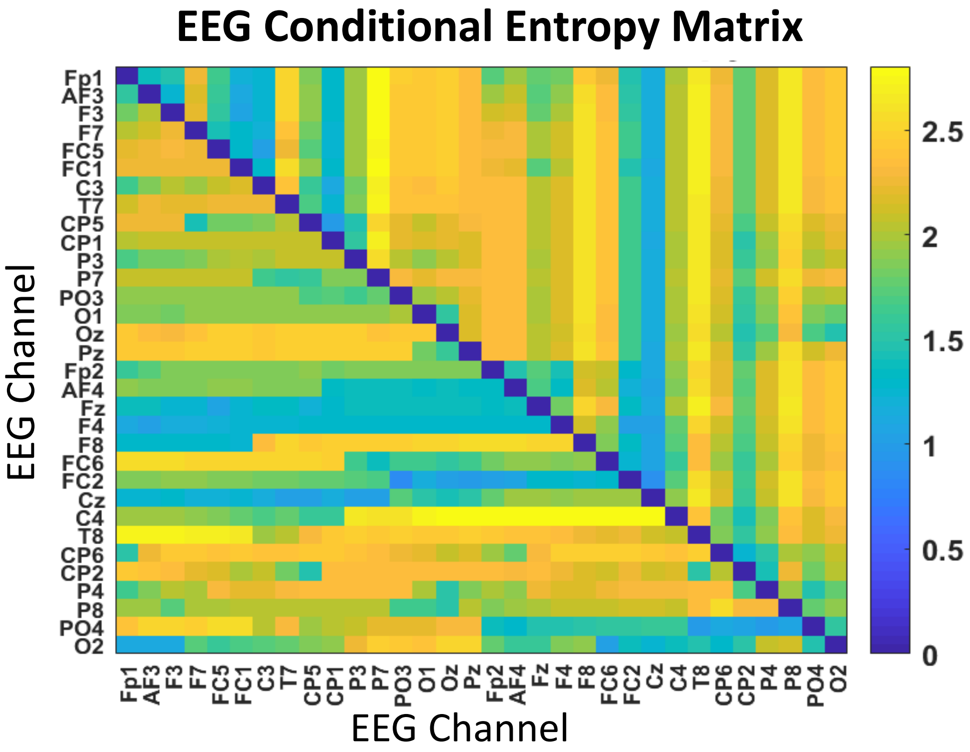
EEG Analysis

- Conditional **entropy** features.
- Used to capture information regarding **interplay** between various brain regions.
- For all possible **pairs** of electrodes.
- 496 features each for DEAP and MAHNOB-HCI datasets and 91 features each for AMIGOS and DREAMER datasets.

Mutual Information: $I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) log \left(\frac{p(x,y)}{p(x)p(y)} \right)$

Conditional entropy H(Y|X): $I(X;Y) = H(Y) - H(Y|X)$

MULTI-MODAL DATA ANALYSIS



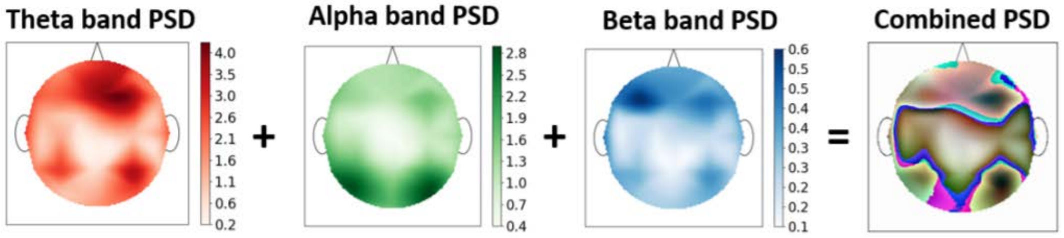
Mutual Information: $I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) log\left(\frac{p(x,y)}{p(x)p(y)}\right)$

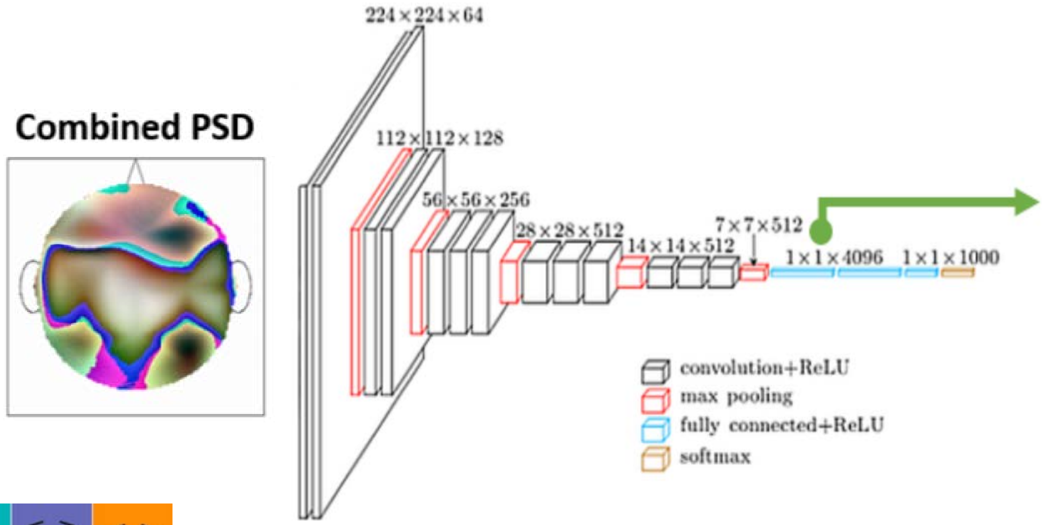
Conditional entropy H(Y|X): $I(X;Y) = H(Y) - H(Y|X)$

EEG Analysis

- Conditional **entropy** features.
- Used to capture information regarding **interplay** between various brain regions.
- For all possible **pairs** of electrodes.
- 496 features each for DEAP and MAHNOB-HCI datasets and 91 features each for AMIGOS and DREAMER datasets.

MULTI-MODAL DATA ANALYSIS





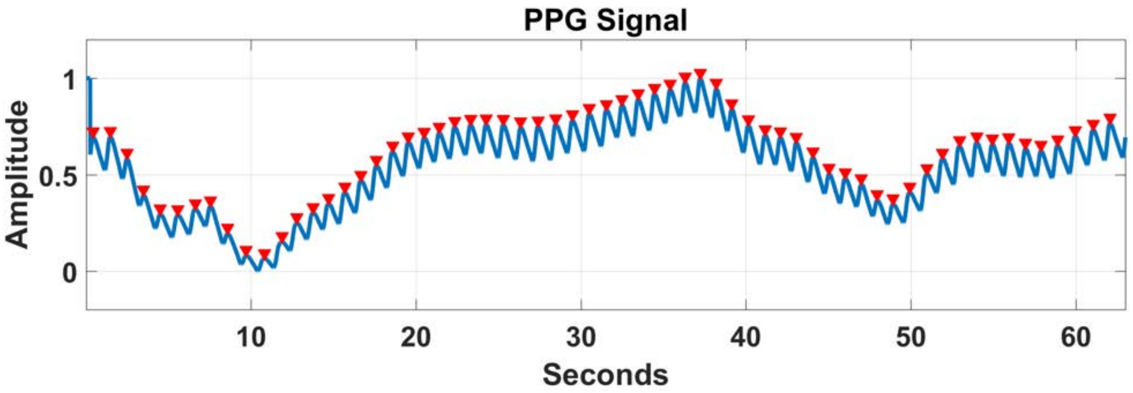
Pre-trained VGG-16 Network

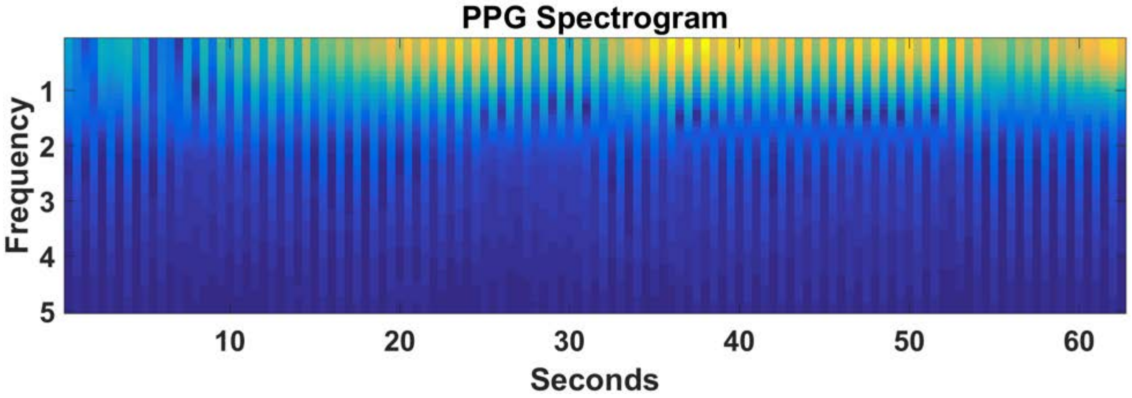
EEG Analysis

- EEG-PSD Deep Learning features.
- **Single image** containing PSD information from the three EEG bands.
- Image is generated **independent** of the number and positions of EEG channels.
- **"Off-the-shelf"** deep learning features from a pre-trained VGG-16 network[1].
- Features from conditional entropy **concatenated** for further analysis.

[1]Simonyan et. al., Very deep convolutional networks for large-scale recognition, *arXiv:1409.1556.*, 2014.

MULTI-MODAL DATA ANALYSIS

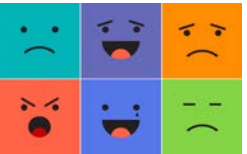




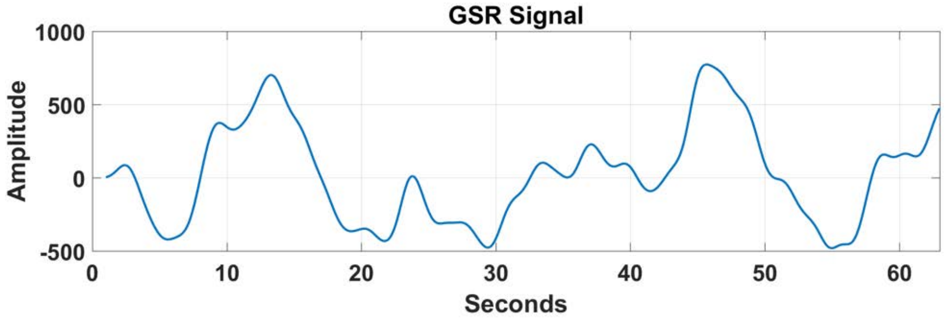
ECG/PPG Analysis

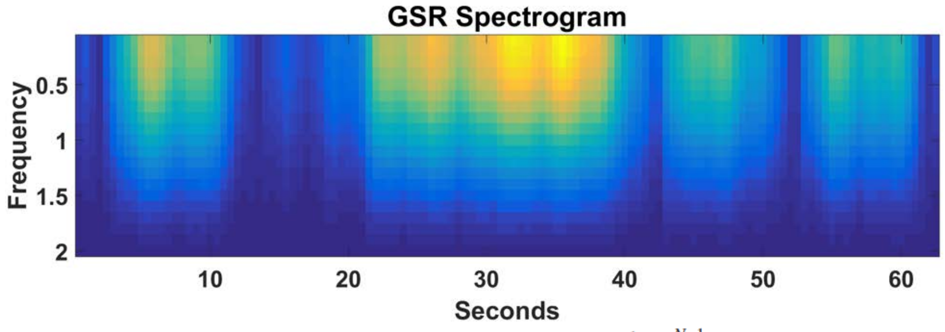
- Low pass filter, **cutoff** @ 60Hz and moving average filter applied to **remove noise.**
- Peaks' **locations** and **heart-rate variability (HRV)** computed.
- Spectrogram computed to extract 4096 **deep learning** features.

Features were calculated for each video (trial) for every subject.



MULTI-MODAL DATA ANALYSIS





$$\tilde{X}_n = \frac{X_n - \mu_X}{\sigma_X}$$

$$\mu_X = \frac{1}{N}\sum_{n=1}^{N} X_n$$

$$\sigma_X = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(X_N - \mu_X)^2}$$

$$\delta_X = \frac{1}{N-1}\sum_{n=1}^{N-1}\left|X_{n+1} - X_n\right|$$

$$\tilde{\delta}_X = \frac{1}{N-1}\sum_{n=1}^{N-1}\left|\tilde{X}_{n+1} - \tilde{X}_n\right|$$

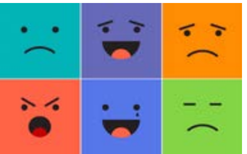
$$\gamma_X = \frac{1}{N-2}\sum_{n=1}^{N-2}\left|X_{n+2} - X_n\right|$$

$$\tilde{\gamma}_X = \frac{1}{N-2}\sum_{n=1}^{N-2}\left|\tilde{X}_{n+2} - \tilde{X}_n\right| = \frac{\gamma_X}{\sigma_X}$$

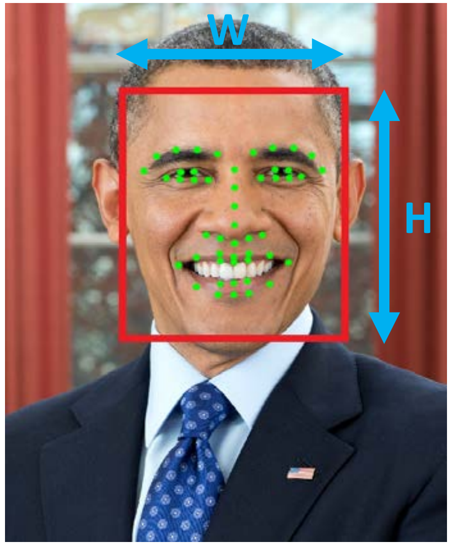
GSR Analysis

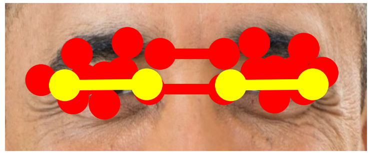
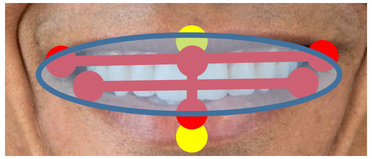
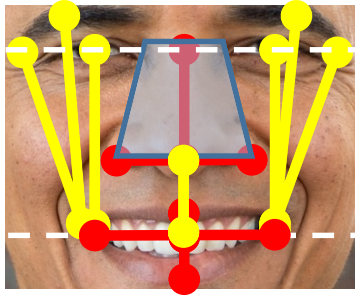
- Low pass filter, **cutoff** @ 60Hz applied and band-pass filter (0.05-1 Hz) applied.
- Peaks' **locations** were computed.
- 8 GSR features based on peaks and nth order moments computed.
- Spectrogram computed to extract 4096 **deep learning** features.

GSR features were calculated for each video (trial) for every subject.



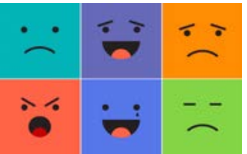
MULTI-MODAL DATA ANALYSIS

Face video analysis (Face – 1)

- One frame extracted for every second.
- Face localization points calculated using **Chehra[1].** Chehra gives 49 face localized points (marked in green).
- **30 features** extracted from localized points based on distances, intersections, angles etc. all normalized over the size of face.
- Some features are the same as calculated for **Action Units[2]** (AU) for emotion recognition.
- Mean, 95th percentile and std. of the above features calculated over all frames in a video (trial).
- 30 features x 3 (mean, median, std) = **90 features**

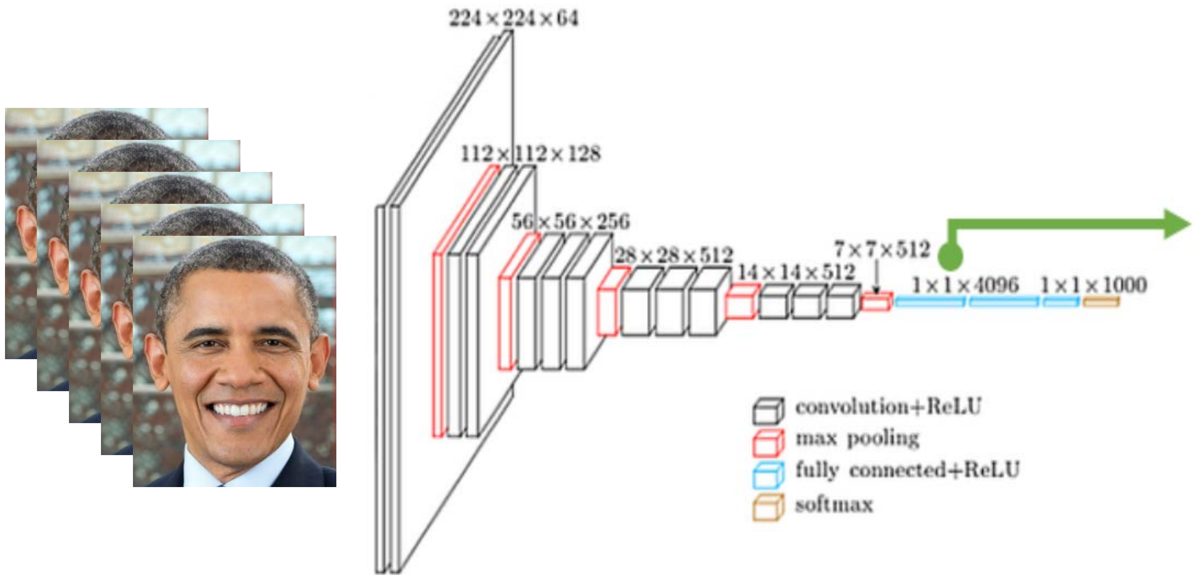


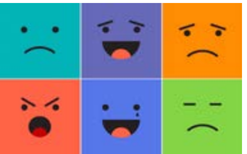
[1]Asthana et. al., Incremental face alignment in the wild, *IEEE CVPR*, 2014.
[2]*Kanade* et. al., Recognizing action units for facial expression analysis, *IEEE Transactions on PAMI*, 2001.

MULTI-MODAL DATA ANALYSIS

Face video analysis (Face – 2)

- Deep Learning features.
- 4096 features extracted using **VGG-Faces** network trained on more than 2.6M images from 2600+ faces[1].
- **Mean, 95th percentile, and std.** of the above features calculated over all frames in a video (trial).





[1]Parkhi et al., Deep face recognition, *British Machine Vision Conference*, 2015.

MULTI-MODAL DATA ANALYSIS

ASL: Average Shot Length

Video Features

- **Shot duration (2 features)**

A measure of the **perceived passage of time.** Can be manipulated by editing effects like cuts, which define the shot length. Also, the number of shots.

- **Visual Excitement**

The **arousal** arising from **motion** in the video.

- **Lighting Key (2 features)**

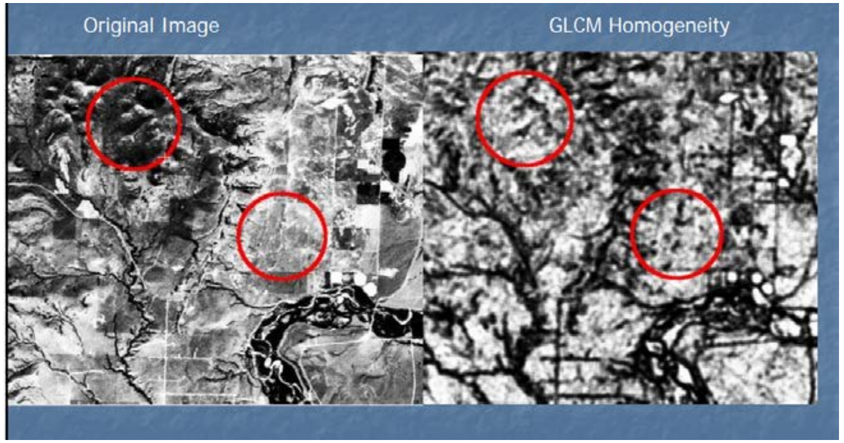
Contrast between light and shadow areas as median and proportion of a frame.

- **Color Energy**

Saturation, brightness and area occupied by **colors.**

Wang et al., Affective understanding in film, *IEEE Transactions on circuits and systems for video technology*, *16*(6), pp. 689-704, 2006.

MULTI-MODAL DATA ANALYSIS



Video Features

- **Grey level co-occurrence matrix (GLCM) features**
  The **distribution** of co-occurring values at a given offset.

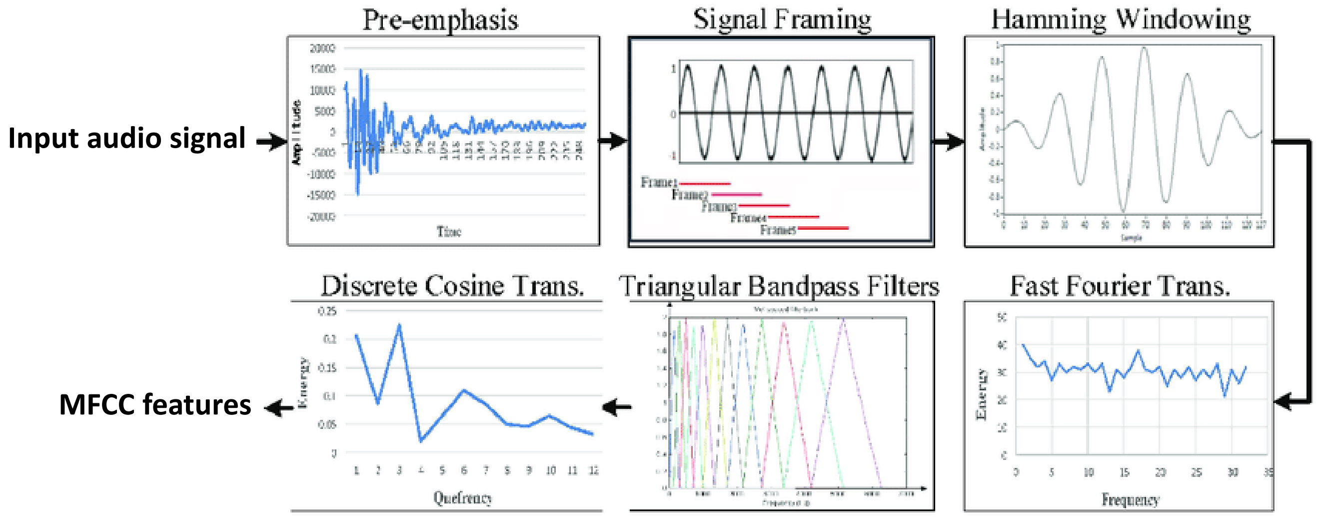
These features **represent** the distance and angular spatial relationship over an image sub-region of a specific size.

Five statistics computed from the GLCM matrix. These provide information about the **texture** of an image:

a) Contrast
b) Correlation
c) Energy
d) Homogeneity
e) Proportion of saturation

- **Total:** 11 video features

MULTI-MODAL DATA ANALYSIS



Audio Features

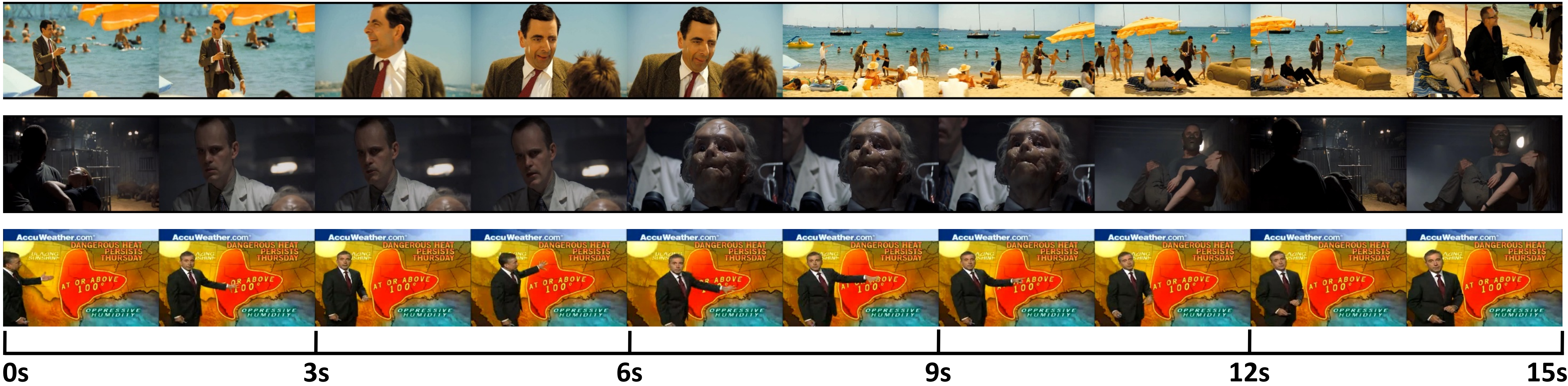
- MFCC Features (13 features)

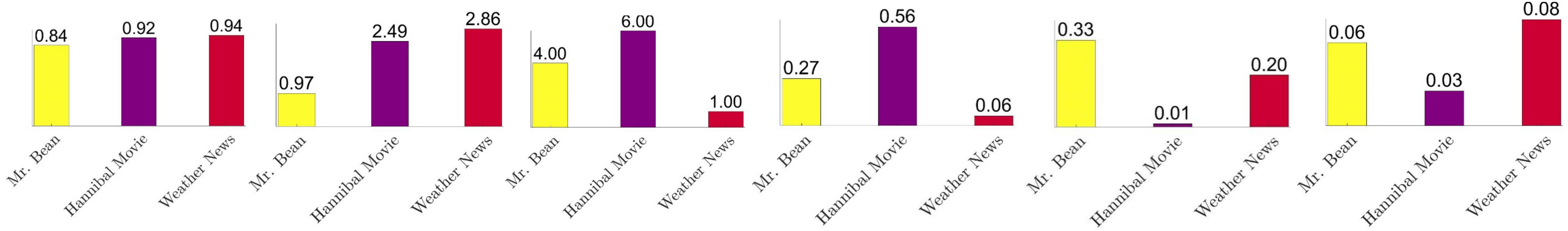
Mel frequency cepstral coefficients. These features model **human perception sensitivity** with respect to frequencies.

- **Loudness** and range of loudness (2 features).
- **Probability** of voice in the sound
- **Tonal** features: Key clarity, mode, and hcdf
- **Total:** 19 audio features

Wang et al., Affective understanding in film, *IEEE Transactions on circuits and systems for video technology*, *16*(6), pp. 689-704, 2006.

Audio-Visual Features Example





Voice Probability

Loudness Range (LU)

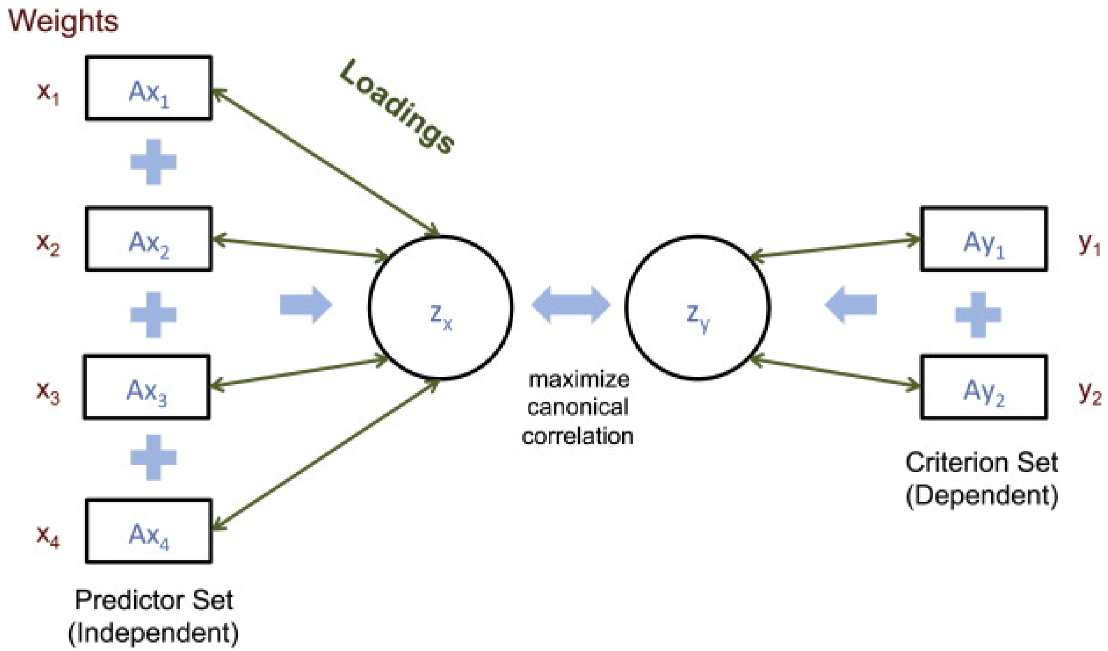
Number of Shots

Proportion of Saturation

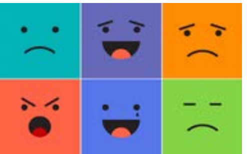
Lighting Key

Texture: Contrast

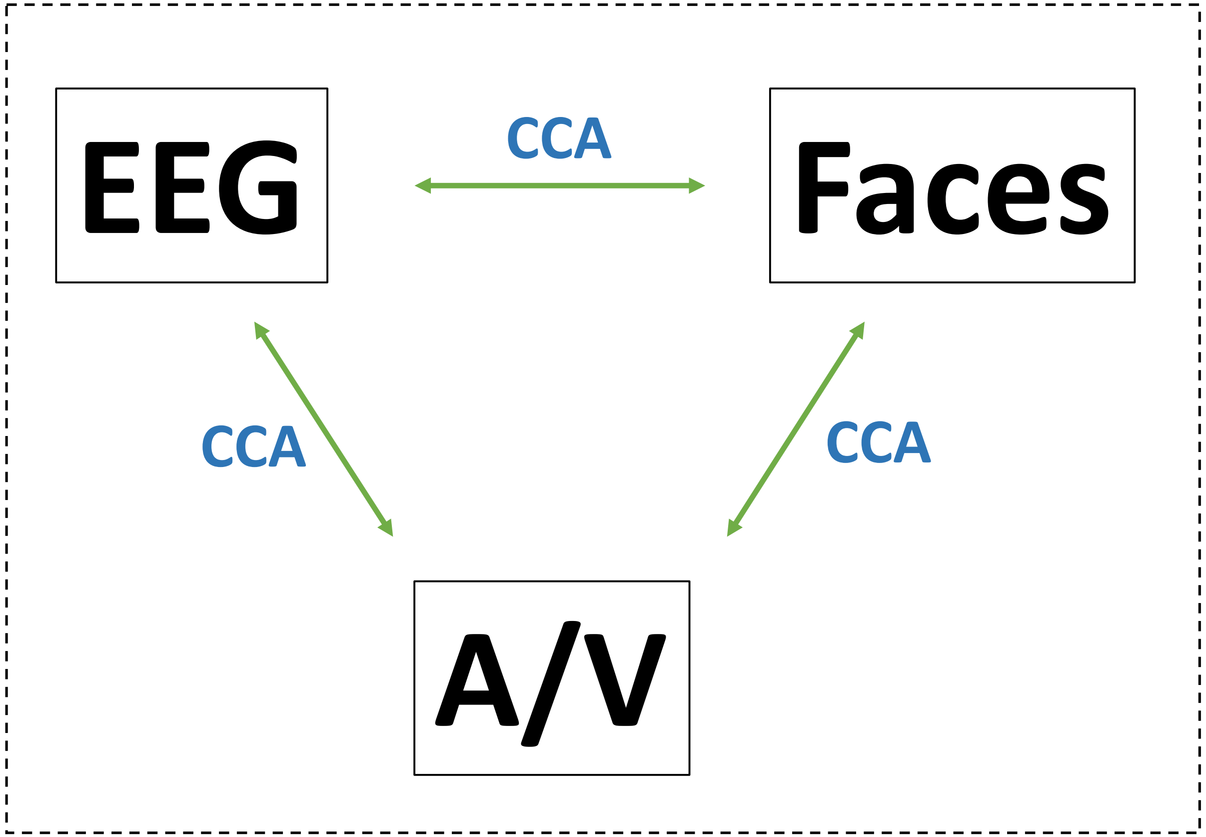
CANONICAL CORRELATION ANALYSIS



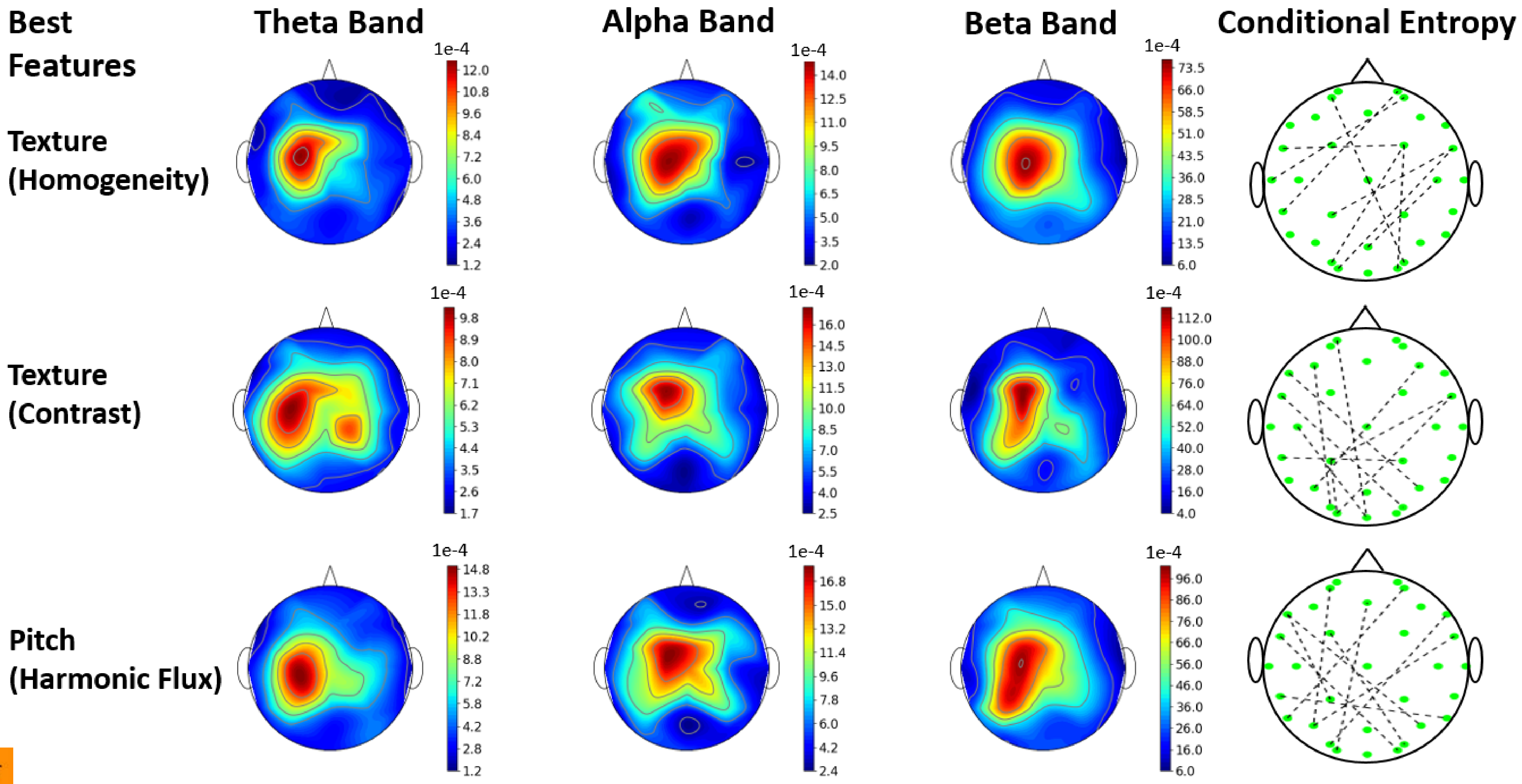
- **15-second** sliding window across all videos (trials) and EEG recordings for all Subjects from the MAHNOB-HCI Dataset. (**> 34,000** total trials)
- Canonical Correlation Analysis **(CCA)** done on the above for each subject **separately.**
- 96 features from the EEG **correlated** with 30 audio-visual features.



CANONICAL CORRELATION ANALYSIS



CCA Between EEG and Audio-Visual Features

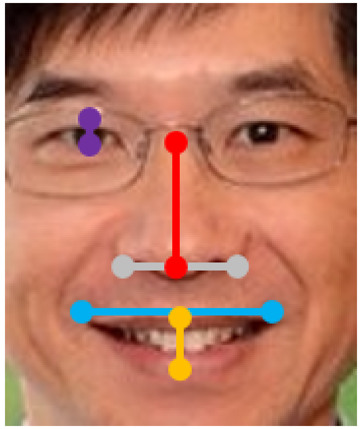


CCA Between Face and Audio-Visual Features

Texture (Homogeneity)

- Nose Area: 0.19
- Lip Height: 0.10
- Eye Height: 0.07
- Nose Height: 0.06
- Nose Width: 0.06

Texture (Contrast)



- Nose Height: 0.12
- Lip Height: 0.08
- Eye Height: 0.08
- Lip Width: 0.07
- Nose Width: 0.06

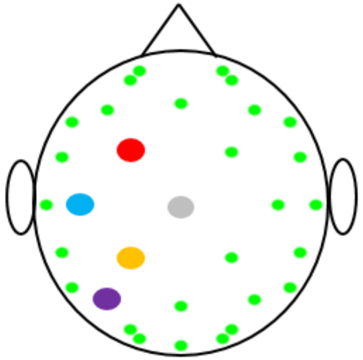
Pitch (Harmonic Flux)

- Nose Area: 0.16
- Eye Height: 0.09
- Lip Width: 0.08
- Nose Height: 0.07
- Nose Width: 0.06

CCA Between EEG and Faces

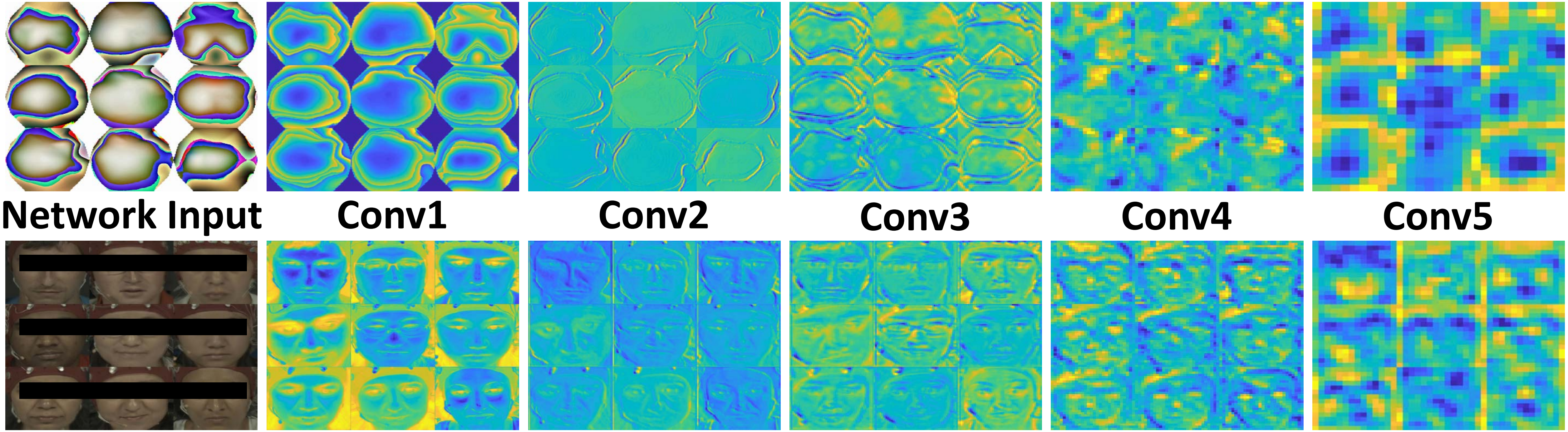
- Top three **EEG** feature maps **across** subjects.

- Nose Width: 51.1%
- Eye Height: 16.2%
- Lip Width: 14.8%
- Lip Height: 6.5%
- Eye Height: 4.1%



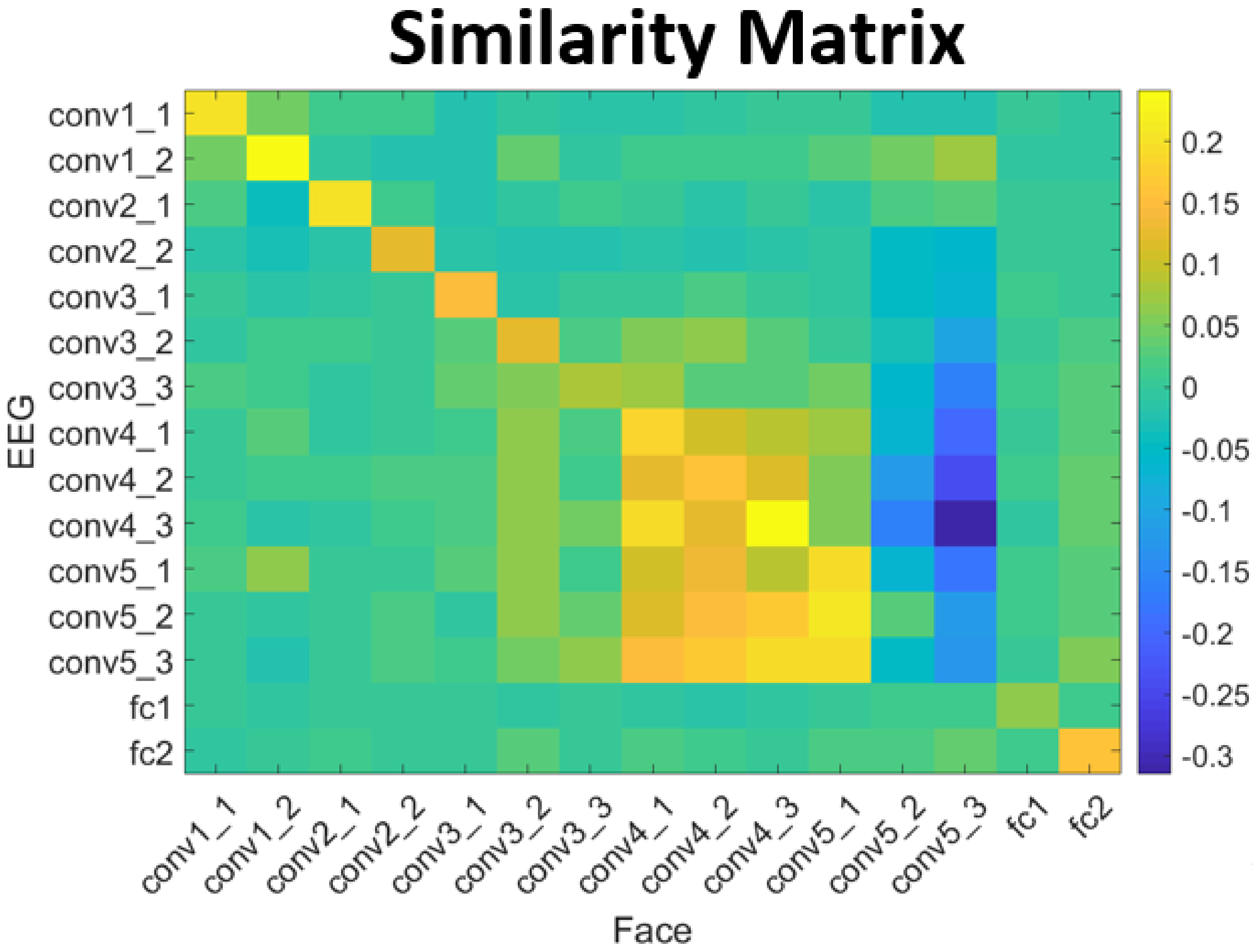
- Beta-FC1: 20.5%
- Beta-CP1: 20.5%
- Beta-P3: 12.2%
- Beta-C3: 11.1%
- Beta-Cz: 7.2%

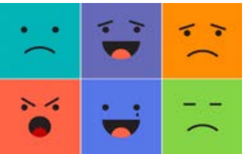
Using VGG-16 Network to Find Correlation



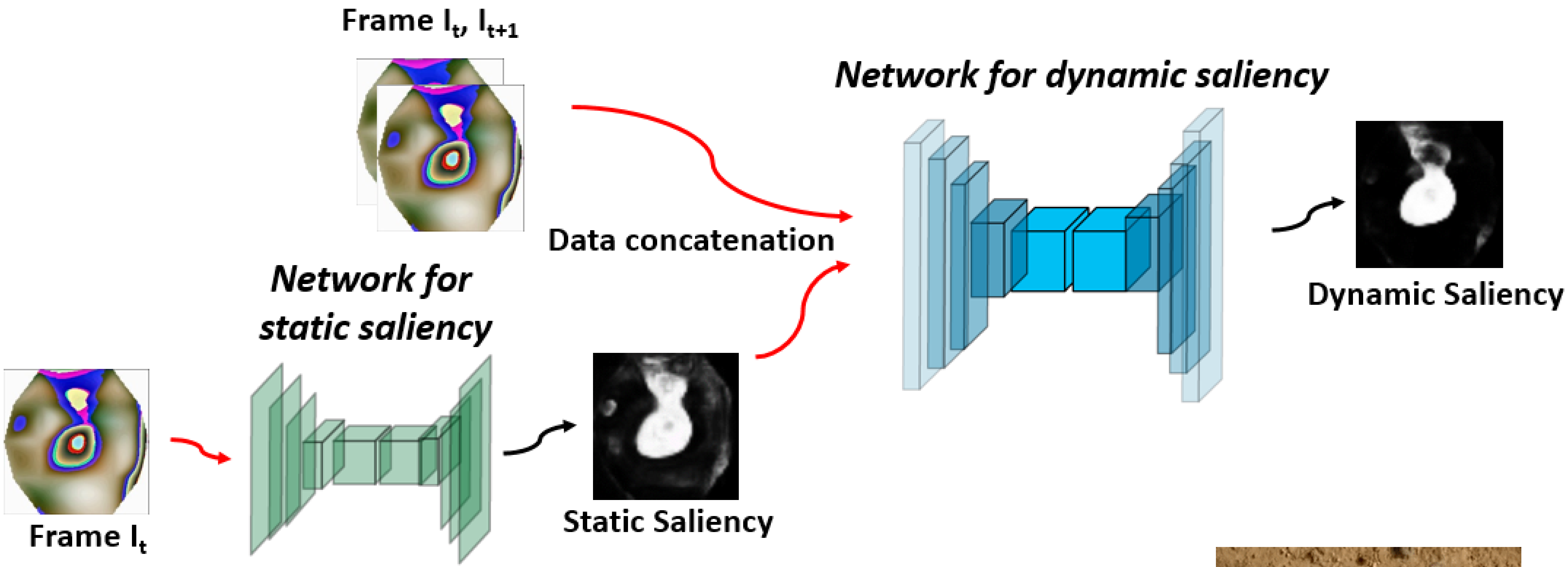
Using VGG-16 Network to Find Correlation

- **Correlation** between **EEG** and **Face** features in deep network:



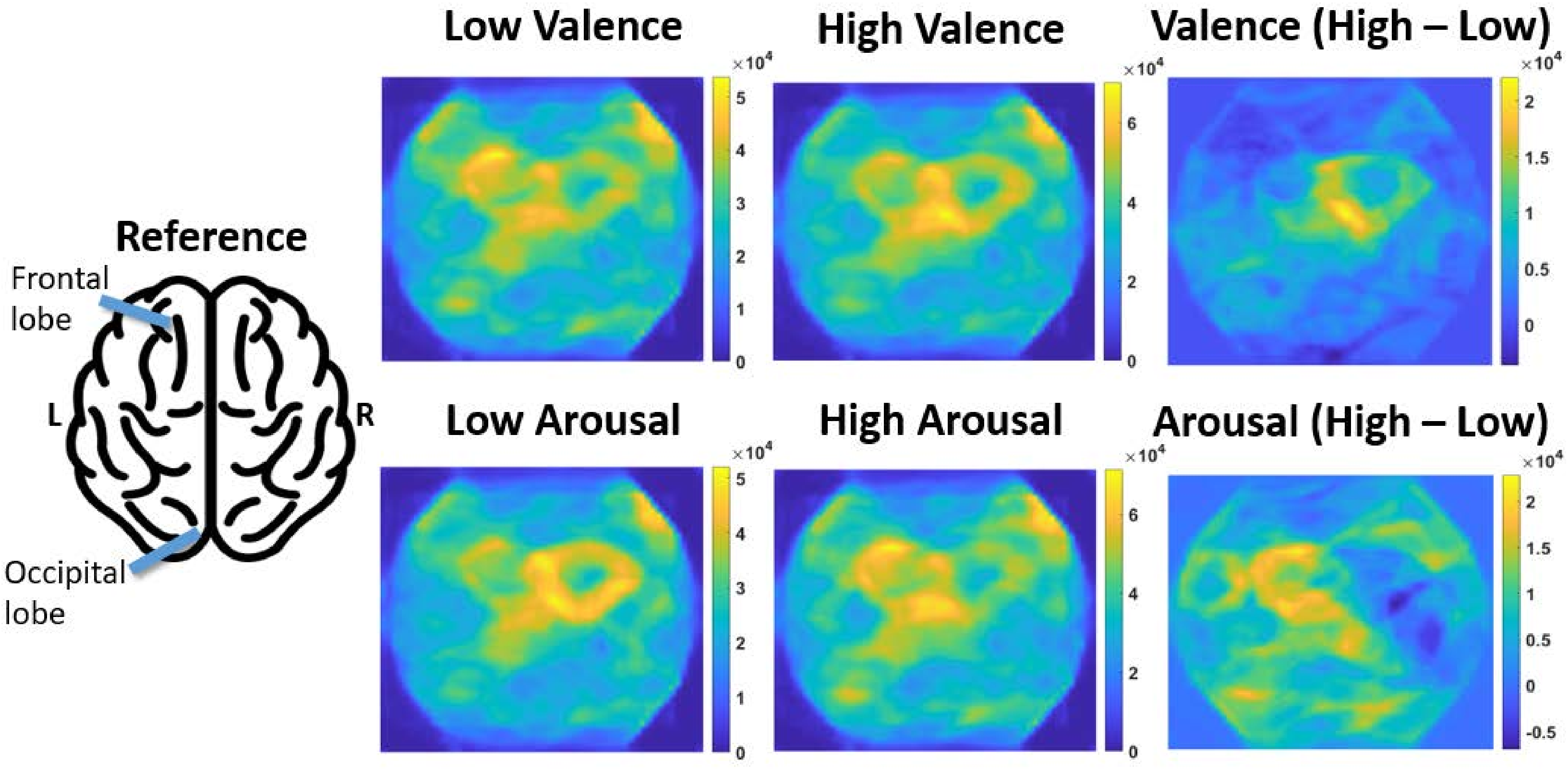


EXTRACTING SALIENT BRAIN REGIONS



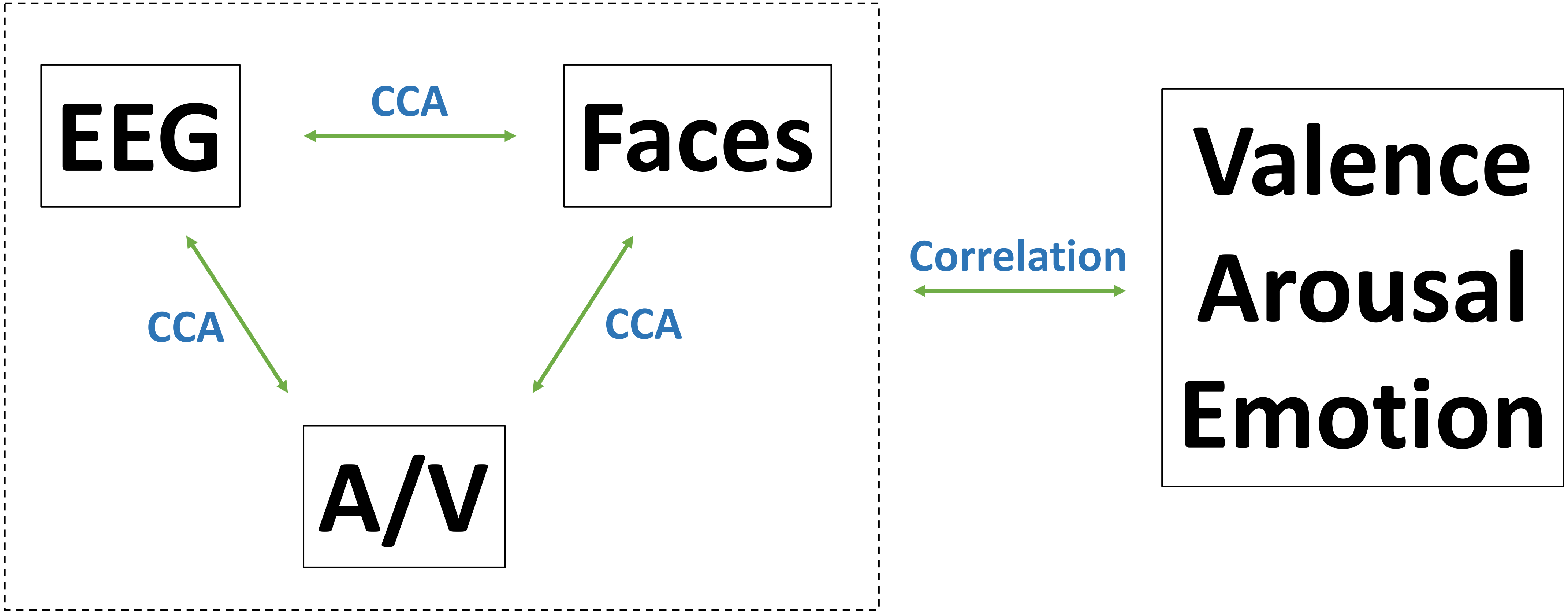
Wang et al., Video salient object detection via fully convolutional networks, *IEEE Transactions on Image Processing*, 2018.

EXTRACTING SALIENT BRAIN REGIONS



An application of **opening** the **deep learning's** Blackbox!

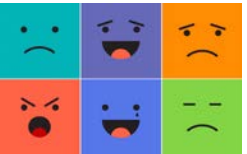
CANONICAL CORRELATION ANALYSIS



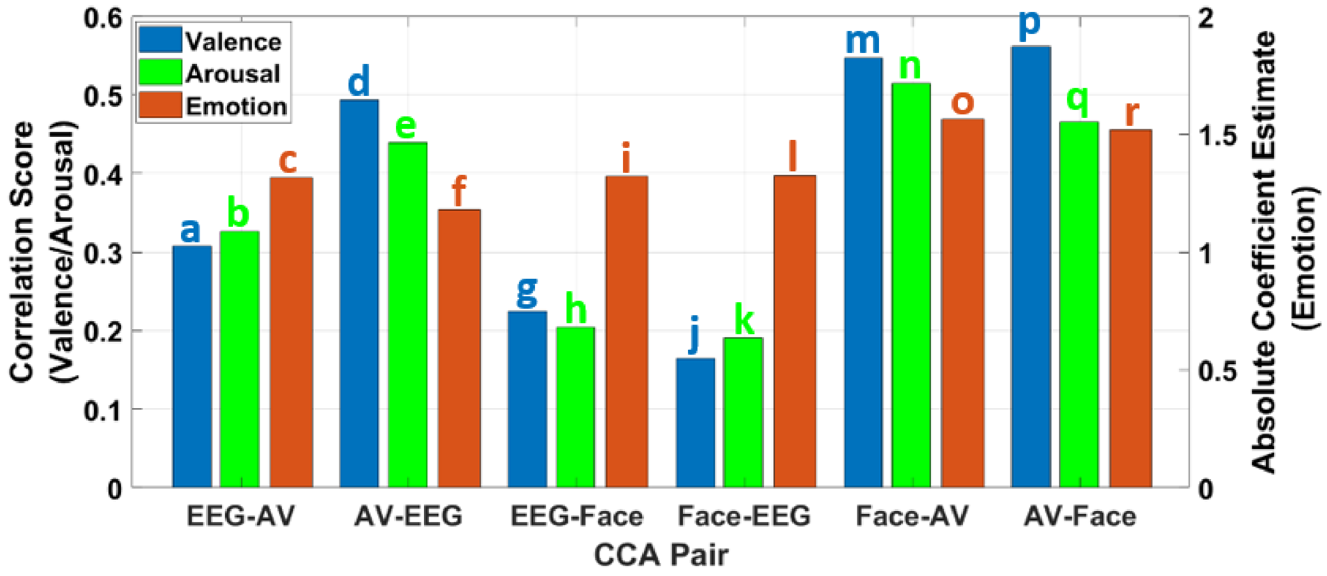
CORRELATION WITH EMOTIONS

- **Valence** distributed between 1 to 9 (integers).
- **Arousal** distributed between 1 to 9 (integers).
- **Emotions** distributed in 12 categories.

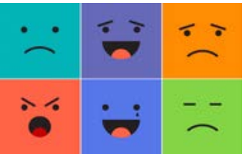
feltEmo#	Emotion name
0	Neutral
1	Anger
2	Disgust
3	Fear
4	Joy, Happiness
5	Sadness
6	Surprise
7	Scream
8	Bored
9	Sleepy
10	Unknown
11	Amusement
12	Anxiety



CORRELATION WITH EMOTIONS



a: Beta-C4	g: Beta-P3	m: Right Eye Height
b: Beta-CP1	h: Beta-Pz	n: Right Eye Height
c: Beta-CP1	i: Beta-FC1	o: Right Eye Height
d: Audio-MFCC 13	j: *d(*right eye, lip*)*	p: Audio-MFCC 13
e: Audio-MFCC 13	k: Right Eye Height	q: Audio-MFCC 13
f: Audio-MFCC 13	l: Right Eye Height	r: Audio-MFCC 13

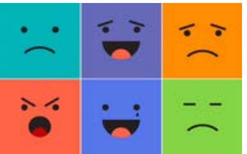


CONTRIBUTIONS

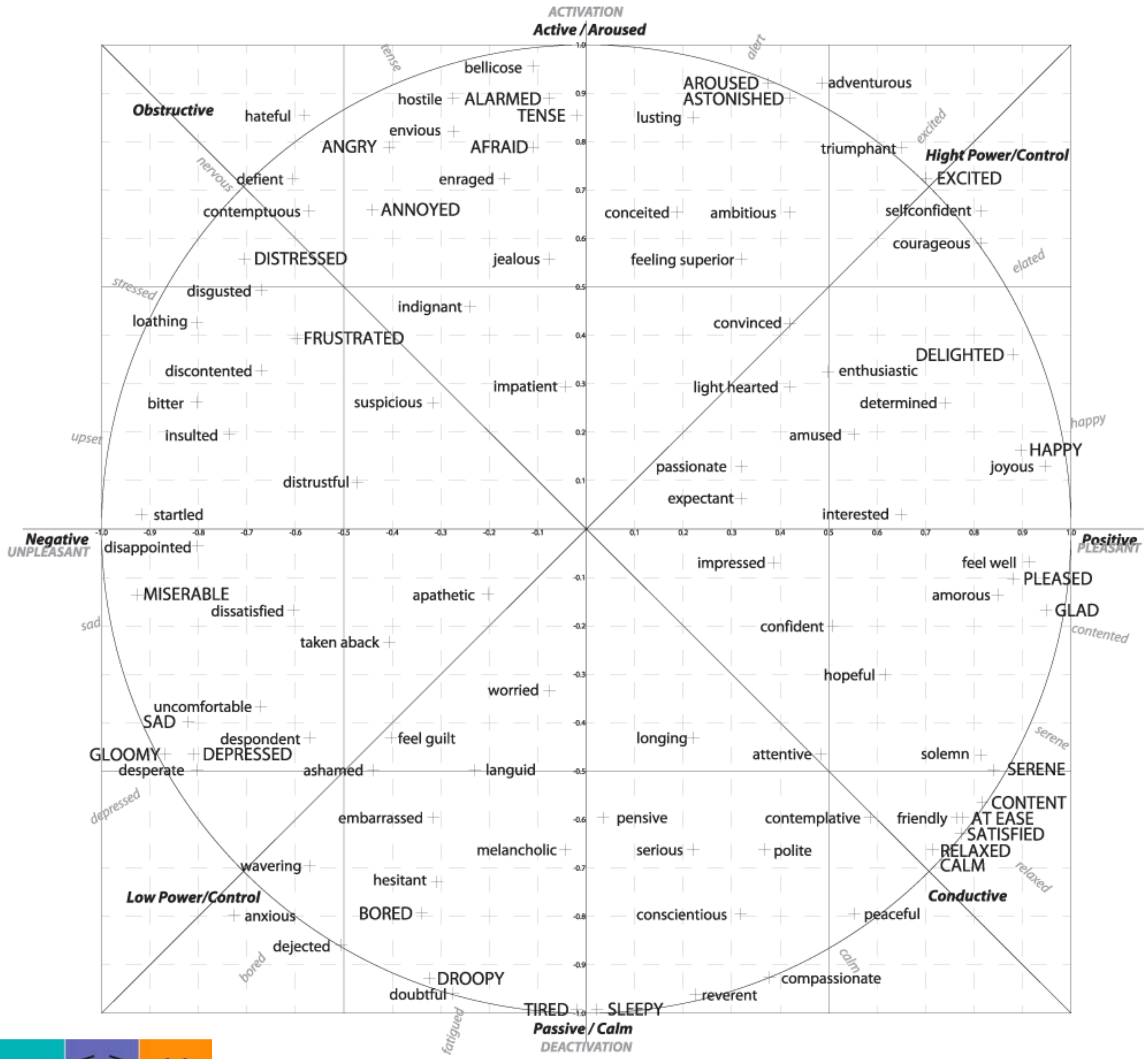
- Represented the features from **two different worlds** i.e. multimedia content and human physiology in the same domain using **CCA.**

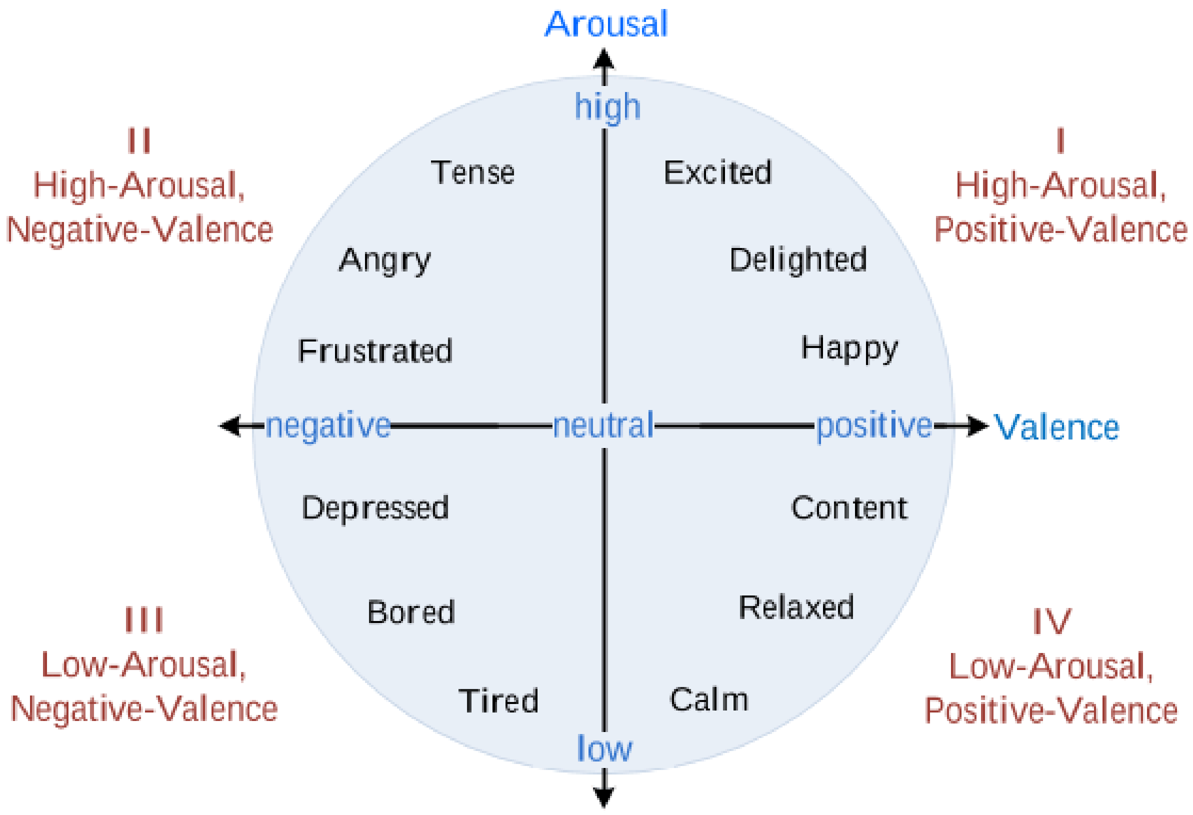
- This **joint analysis** provided insights into which components of the brain EEG and facial expressions **contribute most** toward changes in valence, arousal, and emotions and are **correlated** most with different kinds of multimedia content. In particular, **low-level** features such as texture and color influence human physiology more than **high-level** features such as shot duration, objects, etc.

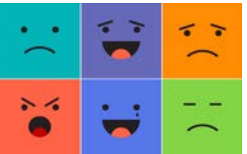
- The **insights** about which audio-visual cues are most **effective** in **evoking** what kind of changes in human physiology. This is useful for designing the **next generation** of **multi-modal** wearables and **bio-sensing** algorithms for use in **affective computing.** These **insights** will also be useful in the domain of **filmmaking.**



AFFECTIVE STATES CLASSIFICATION

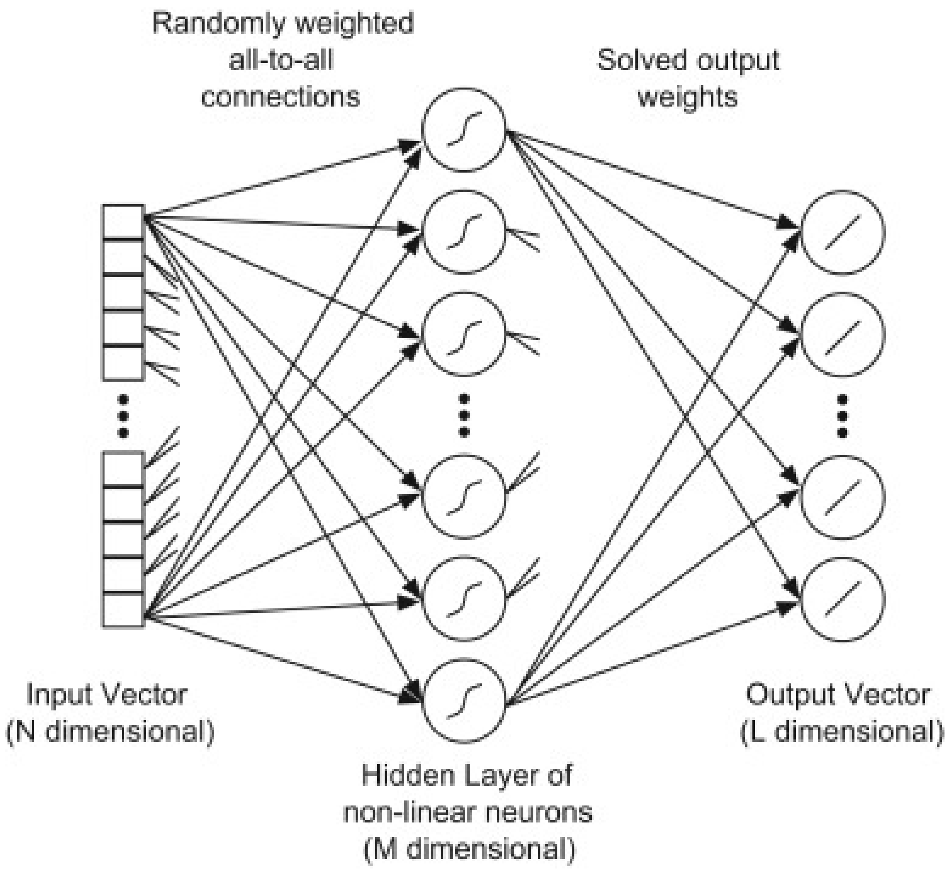






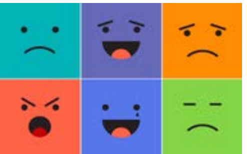
Russell, J.A., A circumplex model of affect, *Journal of personality and social psychology*, 39(6), p. 1161, 1980.

FEATURE CLASSIFICATION



Extreme Learning Machines (ELM) Based Classifier[1]

- Features **re-scaled** between -1 and 1.
- Single **hidden** layer.
- Variable number of neurons.
- Leave-one-subject-out **classification.**
- 10-fold **cross-validation** was performed.
- ELM was chosen since it has been show to work **better** than SVM in previous **affective computing** studies.



[1]Huang et. al., Extreme learning machine: Theory and applications, *Neurocomputing*, 2006.

CLASSIFICATION PERFORMANCE

INDIVIDUAL MODALITY PERFORMANCE EVALUATION

Response	EEG	Cardiac	GSR	Face-1	Face-2
DEAP Dataset					
Valence	71.09/0.68	70.86/0.69	70.70/0.68	71.08/0.68	72.28/0.70
Arousal	72.58/0.65	71.09/0.63	71.64/0.65	72.21/0.65	74.47/0.68
Liking	74.77/0.65	74.77/0.64	75.23/0.64	75.60/0.62	76.69/0.62
Emotion	48.83/0.26	45.55/0.31	45.94/0.25	43.52/0.28	46.27/0.27
AMIGOS Dataset					
Valence	83.02/0.80	81.89/0.80	80.63/0.79	80.58/0.77	77.28/0.74
Arousal	79.13/0.74	82.74/0.76	80.94/0.74	83.10/0.76	77.28/0.72
Liking	85.27/0.81	82.53/0.77	80.47/0.72	80.27/0.72	79.81/0.72
Emotion	55.71/0.30	58.08/0.36	56.41/0.34	57.74/0.28	56.79/0.27
MAHNOB-HCI Dataset					
Valence	80.77/0.76	78.76/0.73	78.98/0.73	83.04/0.79	85.13/0.82
Arousal	80.42/0.72	78.76/0.74	81.84/0.75	82.15/0.77	81.57/0.76
Emotion	57.86/0.33	57.23/0.35	57.84/0.32	60.41/0.35	63.42/0.35
DREAMER Dataset					
Valence	78.99/0.75	80.43/0.78	—	—	—
Arousal	79.23/0.77	80.68/0.77	—	—	—
Emotion	54.83/0.33	57.73/0.36	—	—	—

Denotes mean accuracy/mean F1-score
Number of classes: Valence/Arousal/Liking - 2, Emotion - 4

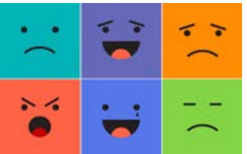
Previous best results

Valence: 76.17% Arousal: 77.19%
Yin et. al., 2017

Valence: 0.58 Arousal: 0.59 (mean F1-score)
Miranda et. al., 2017

Valence: 73% Arousal: 68.5%
Koelstra et. al., 2013

Valence: 62.49% Arousal: 62.32%
Stamos et. al., 2018



CLASSIFICATION PERFORMANCE

MULTI-MODALITY PERFORMANCE EVALUATION

Response	Bio-sensing	EEG and Face	EEG and Face (LSTM)	Previous Best Accuracy
DEAP Dataset				
Valence	71.87/0.68	73.94/0.69	79.52/0.70	77.19
Arousal	73.05/0.68	74.13/0.66	78.34/0.69	76.17
Liking	75.86/0.69	76.74/0.63	80.95/0.70	68.40
Emotion	49.53/0.27	48.11/0.28	54.22/0.31	50.80
AMIGOS Dataset				
Valence	83.94/0.82	78.23/0.74	—	—
Arousal	82.76/0.76	81.47/0.72	—	—
Liking	83.53/0.77	81.49/0.75	—	—
Emotion	58.56/0.40	58.02/0.29	—	—
MAHNOB-HCI Dataset				
Valence	80.36/0.75	85.49/0.82	—	73.00
Arousal	80.61/0.71	82.93/0.77	—	68.50
Emotion	58.07/0.30	62.07/0.35	—	—
DREAMER Dataset				
Valence	79.95/0.77	—	—	62.49
Arousal	79.95/0.77	—	—	62.32
Emotion	55.56/0.33	—	—	—

Denotes mean accuracy/mean F1-score

Number of classes: Valence/Arousal/Liking - 2, Emotion - 4

Previous best results

Valence: 76.17% Arousal: 77.19%
Yin et. al., 2017

Valence: 0.58 Arousal: 0.59 (mean F1-score)
Miranda et. al., 2017

Valence: 73% Arousal: 68.5%
Koelstra et. al., 2013

Valence: 62.49% Arousal: 62.32%
Stamos et. al., 2018

CLASSIFICATION PERFORMANCE

COMBINED DATASET PERFORMANCE EVALUATION

Response	EEG	Cardiac	GSR	Face-1	Face-2
DEAP + AMIGOS Combined Dataset					
Valence	62.80/0.58	59.69/0.59	59.64/0.58	63.04/0.62	62.38/0.62
Arousal	62.27/0.61	63.61/0.61	61.98/0.62	67.66/0.65	68.65/0.66
Liking	69.13/0.59	69.27/0.61	69.27/0.55	67.99/0.64	68.65/0.64
Emotion	37.47/0.27	37.50/0.22	37.24/0.31	40.92/0.36	42.24/0.36
DEAP + AMIGOS + MAHNOB-HCI Combined Dataset					
Valence	61.24/0.60	58.57/0.59	58.98/0.57	61.59/0.61	62.56/0.63
Arousal	65.15/0.63	61.84/0.61	61.02/0.59	65.94/0.65	67.15/0.66
Emotion	40.21/0.35	36.33/0.31	35.71/0.28	42.51/0.33	43.00/0.32

TRANSFER LEARNING PERFORMANCE EVALUATION

Response	EEG	Cardiac	GSR	Face-1	Face-2
DEAP + AMIGOS (Train Dataset), MAHNOB-HCI (Test Dataset)					
Valence	63.55/0.60	64.77/0.54	64.96/0.55	55.02/0.52	62.01/0.62
Arousal	58.37/0.55	62.50/0.52	62.50/0.52	59.32/0.54	58.60/0.58
Emotion	36.65/0.32	39.58/0.28	38.64/0.28	36.38/0.39	34.05/0.37
DEAP (Train Dataset), MAHNOB-HCI (Test Dataset)					
Valence	62.70/0.54	63.59/0.46	65.19/0.47	56.48/0.49	59.86/0.59
Arousal	61.99/0.55	61.46/0.48	63.23/0.52	59.33/0.56	61.99/0.60
Emotion	35.88/0.23	38.01/0.24	39.08/0.24	33.57/0.33	32.50/0.22

Denotes mean accuracy/mean F1-score
Number of classes: Valence/Arousal/Liking - 2, Emotion - 4

CONTRIBUTIONS

- The **most comprehensive** **affective computing** study to-date utilizing four datasets containing data from 122 subjects and 2800+ trials. We were able to **beat** the previous best results for the four datasets.

- The features were extracted **intuitively** from the four **bio-sensing** modalities (such as mutual information in EEG, face-localized point-based in face tracking, etc.) as well as from the **black-box** deep learning perspective. It was the **fusion** of these features that proved significant in boosting the performance.

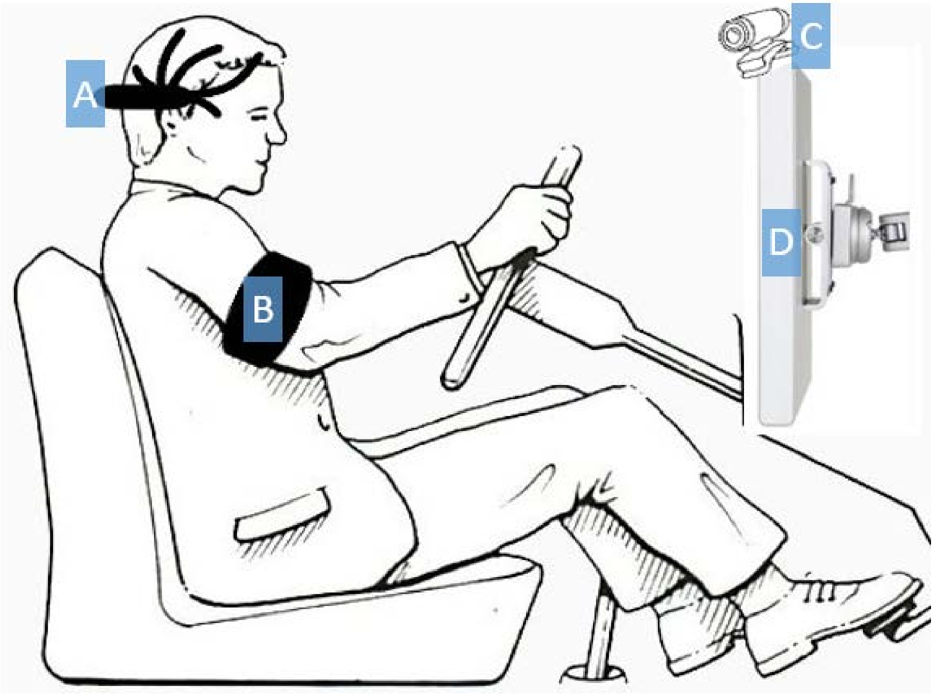
- The features proved to perform well even **across datasets** and **transfer learning** among them (**significantly** above chance accuracy) showing that the choice of features by us was to an extent highly **robust** and **scalable.**

How to apply them toward **Real-world** applications?

- Consuming Multimedia Content
- Monitoring Driver Awareness

DRIVER AWARENESS ANALYSIS

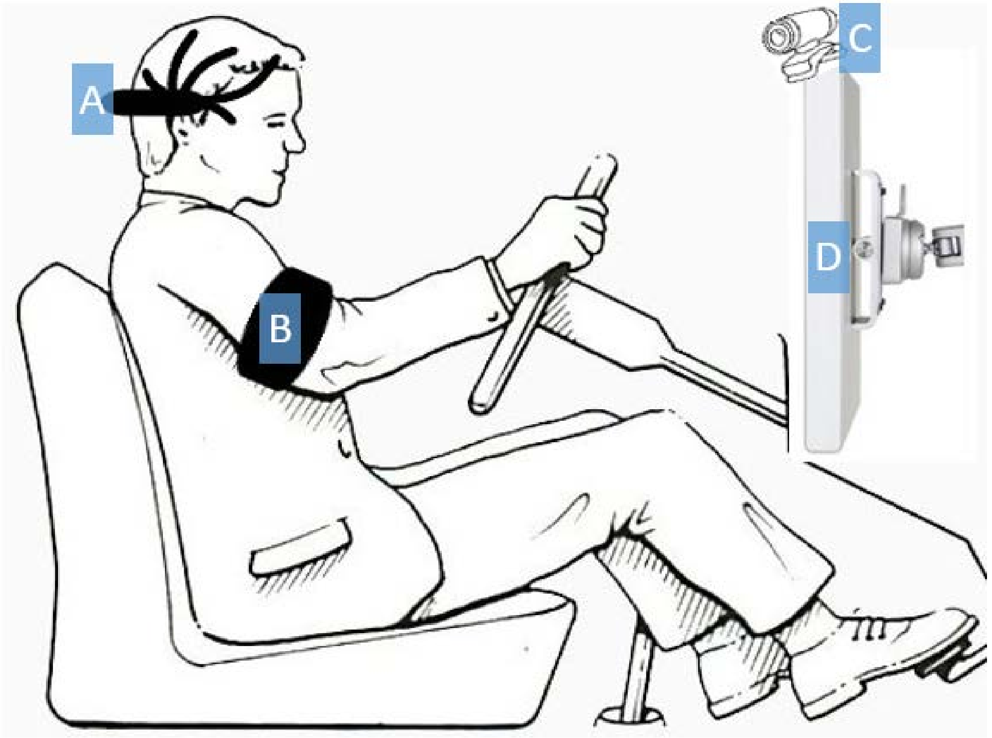
Affective Computing is the study and development of systems and devices that can recognize, interpret, process, and simulate human affects **(feeling, emotion, or mood).**



- **Attention monitoring** is a subfield under **Affective Computing.**
- **Attention monitoring** is **crucial** since one out of five automobile **crashes** happen due to falling **asleep.**[1]
- Driver awareness has a direct correlation with how **attentive** the driver is.
- Goal was to monitor the **driver's attention** during different scenarios such as driving on the freeway, in a narrow street etc.
- Another goal was to **assess the driver's facial and EEG response** towards short-duration hazardous events.

[1]httpps://www.washingtonpost.com/news/dr-gridlock/wp/2014/11/04/falling-aslee-causes-1-in-5-auto-crashes/

DRIVER AWARENESS ANALYSIS



Driving simulator with real-drive videos

- 14-channel **EEG**, **PPG**, **GSR**, and **video camera.**
- 12 participants.
- 35 videos (30-90 seconds long)
- 15 videos from public **KITTI Dataset**[1] and 20 videos collected around San Diego using **LISA-T** (Tesla Model S) vehicle. KITTI Dataset contains videos from Karlsruhe, Germany.
- KITTI Dataset was used to **compare** the performance with existing research studies (**AUC Performance** with EEG: 0.79)[2].

[1]Geiger et al., Vision meets robotics: The KITTI dataset, The *International Journal of Robotics Research*, 2013. [2]Kolkhorst et al., Decoding hazardous events in driving videos, *7th Graz Brain-Computer Interface Conference*, 2017.

DRIVER AWARENESS ANALYSIS

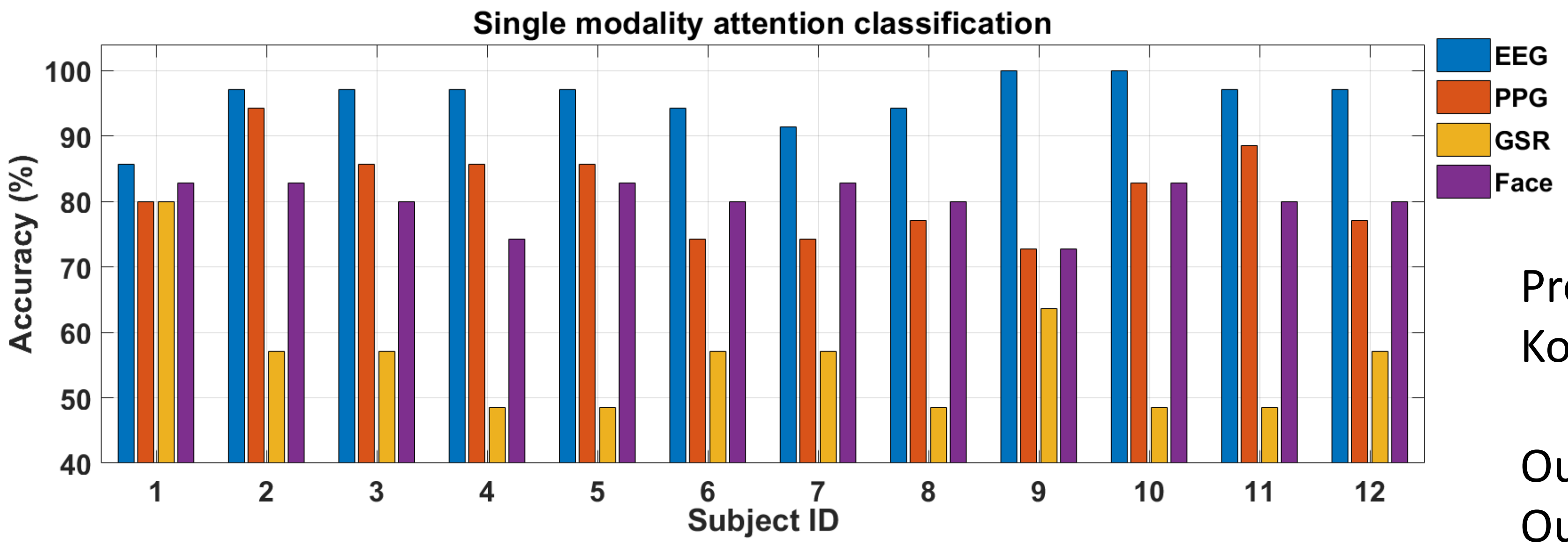
(A)

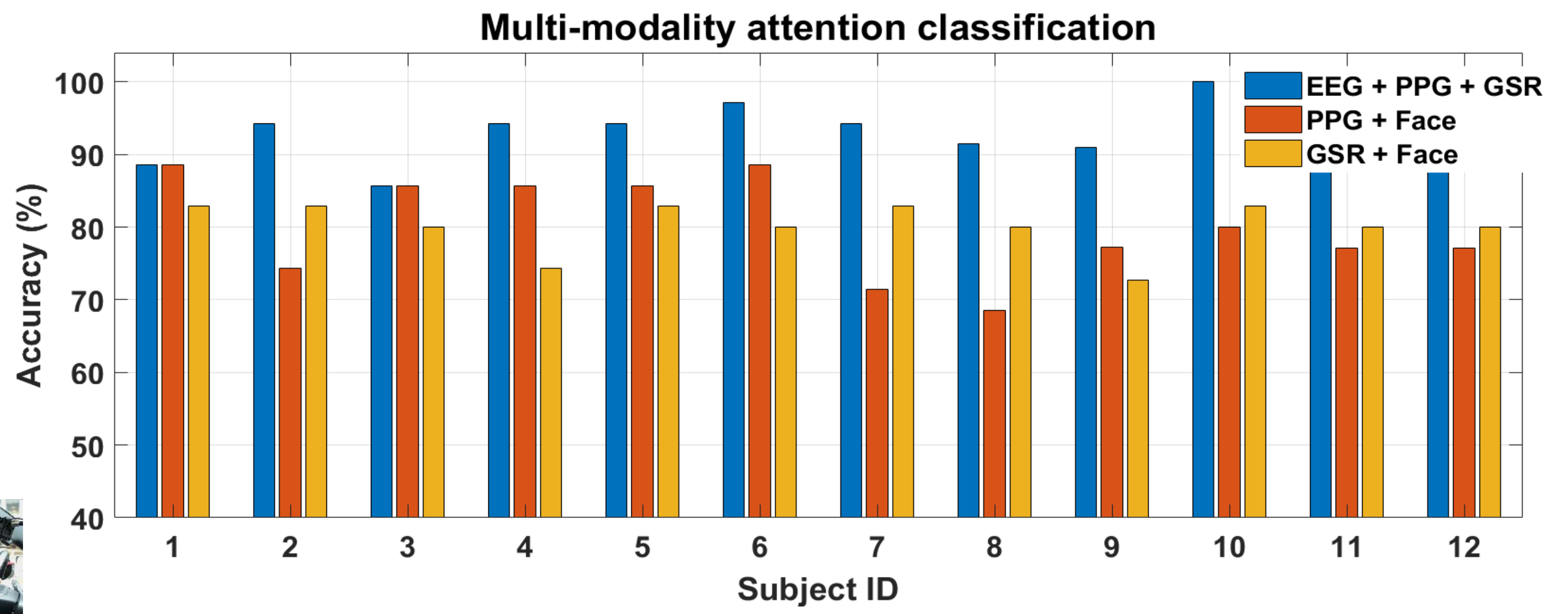
(B)

Various image **instances** from videos collected in (A) LISA Dataset and (B) KITTI Dataset

Previous research studies **only** utilized a **single** dataset and a **single** sensor modality whereas we implement a **multi-moda**l approach to driver **awareness** analysis.

ATTENTION CLASSIFICATION (LOW/HIGH)





Previous **best** results
Kolkhorst et al. EEG AUC: 0.79

Our EEG AUC: 0.84
Our PPG AUC: 0.83
Our GSR AUC: 0.71
Our Face AUC: 0.79

Our EEG + PPG + GSR AUC: 0.85
Our PPG + Face AUC: 0.80
Our GSR + Face AUC: 0.80

HAZARDOUS EVENTS CLASSIFICATION

(A)

(B)

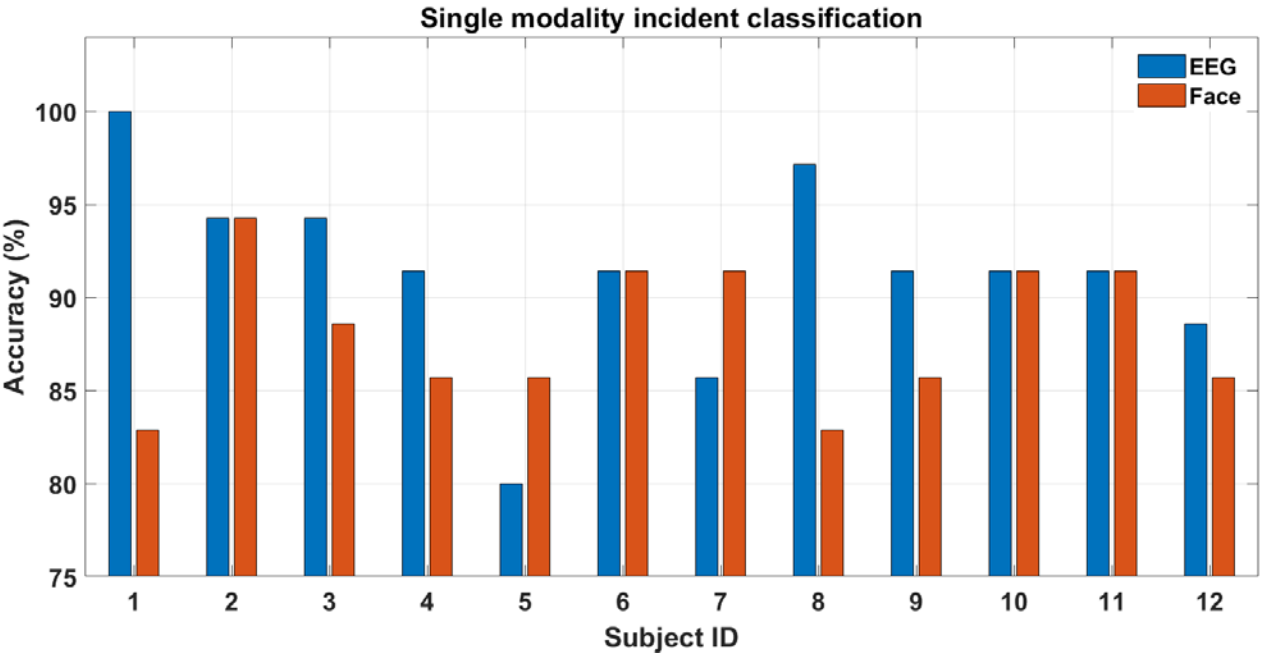
Hazardous/Non-hazardous incident classification

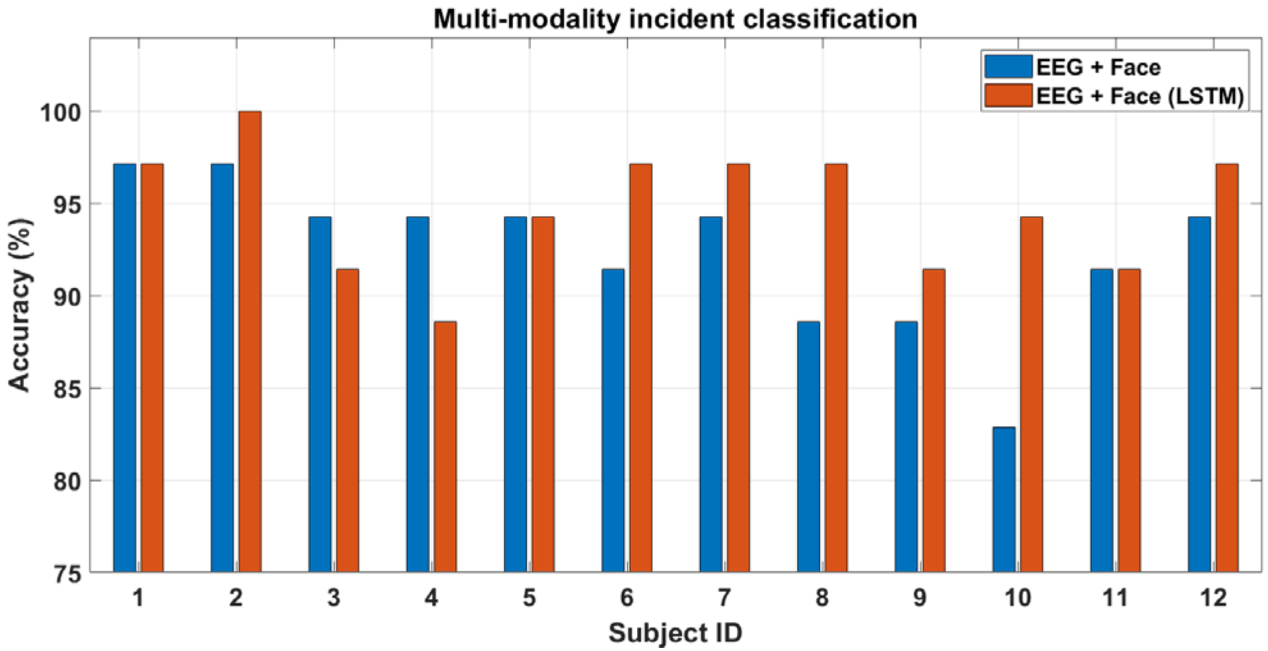
- **2-seconds** of **hazardous/non-hazardous** events marked.
- 30 hazardous and 40 non-hazardous incidents.
- Leave-one-subject-out **cross validation.**

(A) Hazardous incidents
KITTI Dataset (above)
LISA Dataset (below)

(B) Non-hazardous incidents
KITTI Dataset (above)
LISA Dataset (below)

HAZARDOUS EVENTS CLASSIFICATION

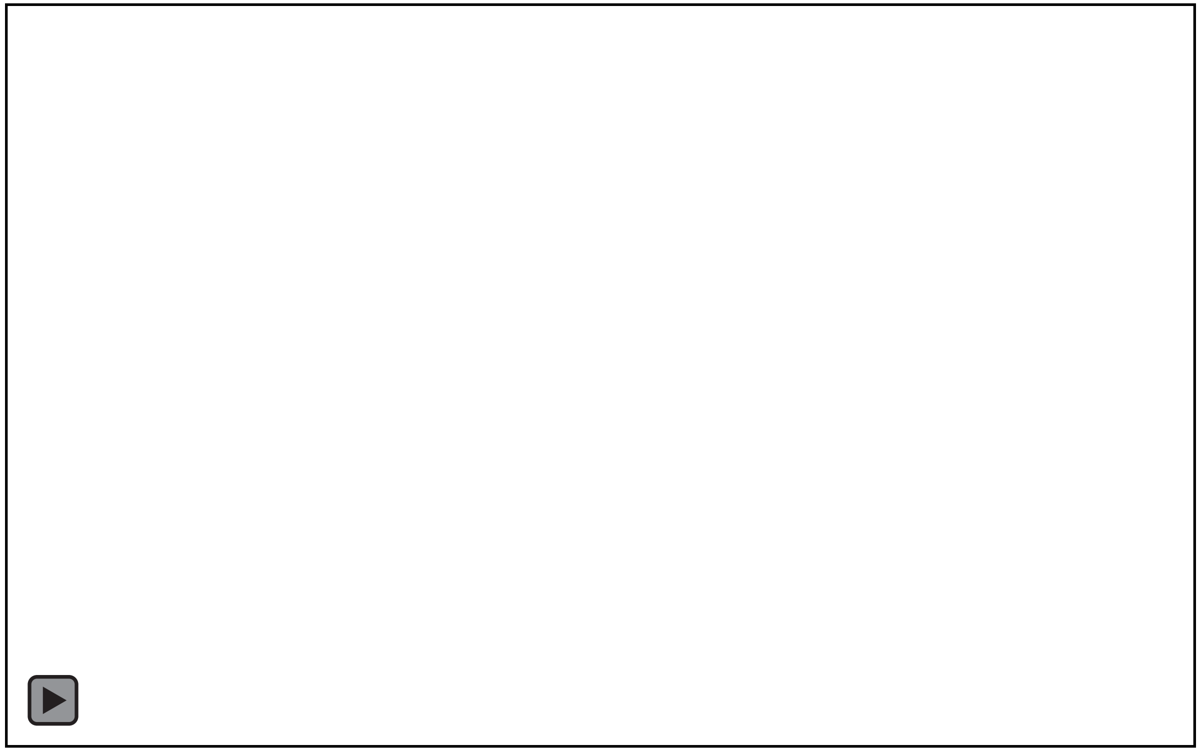




Modality	Attention Analysis	Incident Analysis
EEG	$95.71 \pm 3.95\%$	$91.43 \pm 5.17\%$
Faces	$80.11 \pm 3.39\%$	$88.10 \pm 3.82\%$
EEG + Faces	$95.10 \pm 3.62\%$	$92.38 \pm 4.10\%$
EEG + Faces (LSTM)	—	$\boxed{94.76} \pm 3.41\%$

NOVEL DRIVING + MULTIMEDIA DATASET

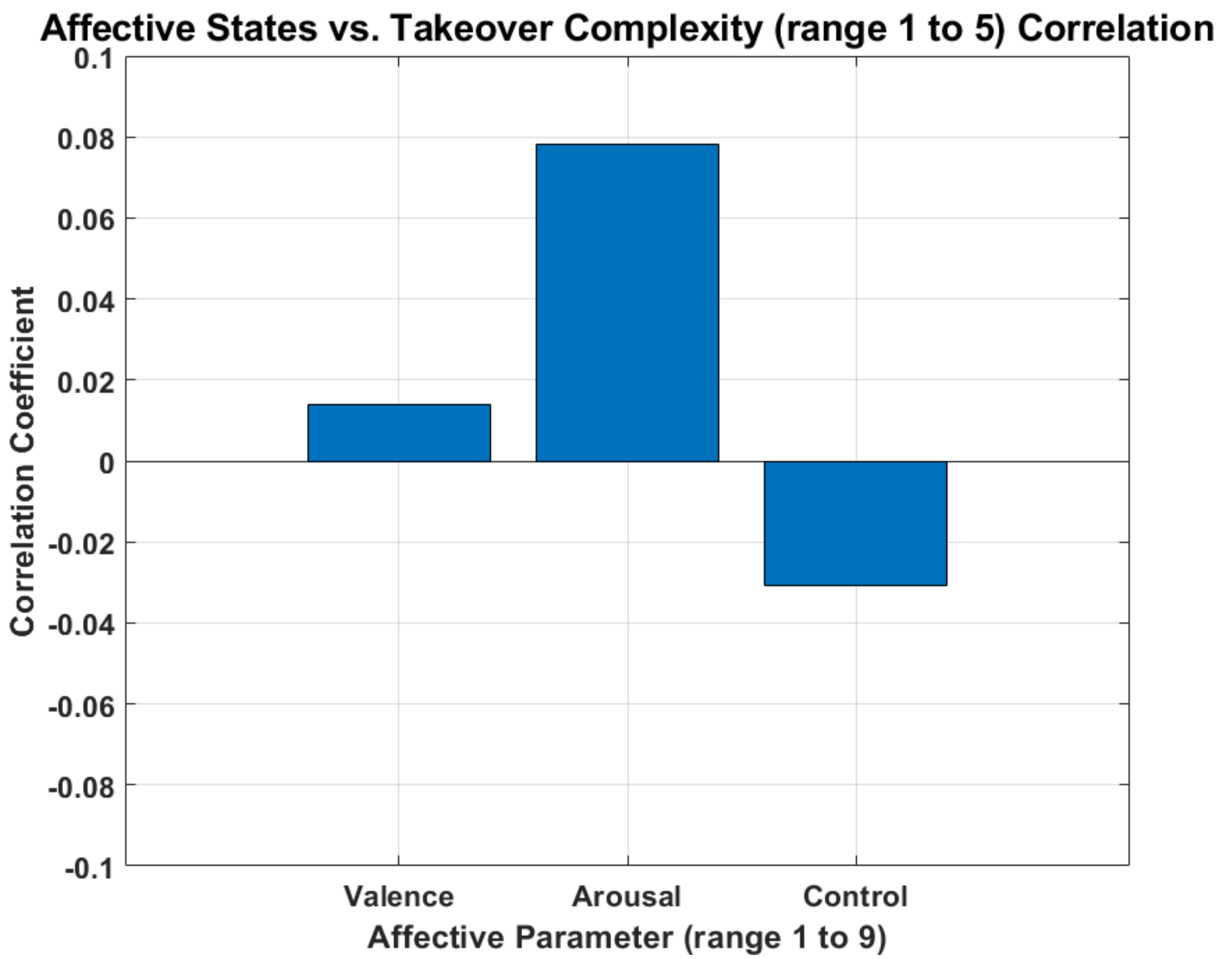
Tesla S Interior

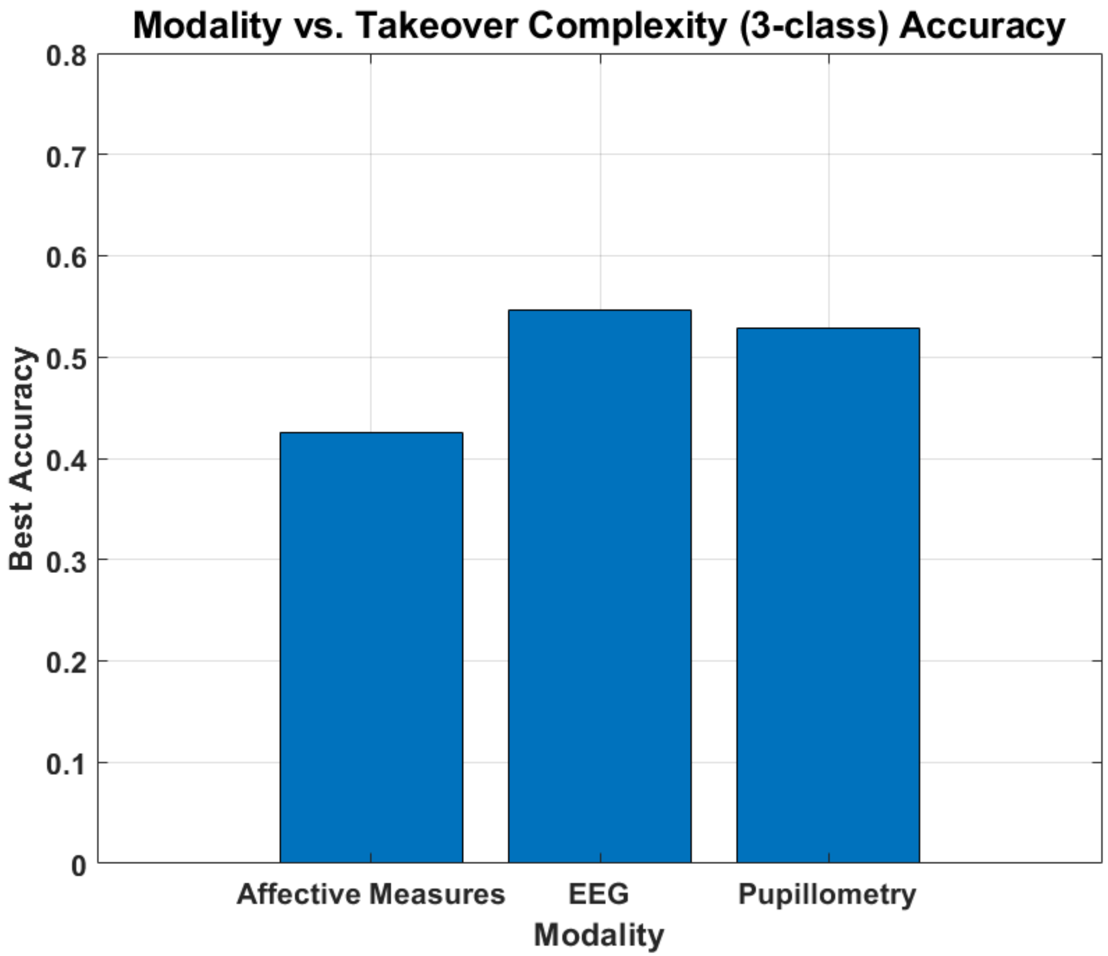


Watching News in Autopilot Mode -> Takeover Beep -> Driving

After each takeover, users rate **takeover complexity** on a scale of 1 (very easy) to 5 (very hard).

DRIVING + MULTIMEDIA RESULTS





EEG and Pupillometry (diameter, fixations, saccades) features calculated over the **last three seconds** just **before takeover.** Linear **SVM** used for classification.

CONTRIBUTIONS

- It was **evaluated** if the modalities with **low-temporal resolution** (but easily **wearable**) namely PPG and GSR can work as well as EEG and vision modality for assessing driver's **attention** and **hazard** analysis. The **outcome** of this hypothesis turned out to be **negative.**

- The **efficacy** of the **fusion** of features from different modalities i.e. using **multi-modal** systems was **evaluated** for **attention** and **hazard** analysis. Again, EEG and vision and their **combination** provided the **best** performance. **Previous** research studies **only** focused on either vision or EEG and no **multi-modal** approaches were reported.

- These **insights** will enable the design of **safer automobiles** and **integrating** their software with **bio-sensing wearable** devices such as Fitbit, Apple Watch, etc. in **addition** to using cabin cameras inside the vehicle.

FIVE Ws and One H

- **Who** – Siddharth and collaborators
- **Where** – UC San Diego and Facebook Reality Labs
- **What** is **Affective Computing**?
- **Why** use **Bio-sensing**?
- **When** are **Multi-modal** tools advantageous?
- **How** to apply them toward **Real-world** applications?

Goals of such a Bio-sensing system

- Detect and monitor **affective** states.
- Infer **affective** states using a **minimal** number of and most **comfortable** sensors.
- Infer the **context** in **real-world** scenarios.
- Make **recommendations**/take action based on the information from above.
- Do all the above **continuously** throughout the day.

Where will this all lead to?

- Detect and monitor **affective** states.

- Infer **affective** states using a **minimal** number of and most **comfortable** sensors.

- Infer the **context** in **real-world** scenarios.

- Make **recommendations**/take action based on the information from above.

- Do all the above **continuously** throughout the day.

CONCLUSION

- **Affective computing** encompasses the development of systems that can work in a multitude of **challenging conditions** since human affects are **highly subjective.** The same person may react differently to multimedia content at different times while different people may react differently to the same content. Herein lies **the need** for recording the user's **physiology.**

- **Multi-modal bio-sensing** systems are our **best bet** for now since no single modality can **efficiently** capture human affects **continuously** under **real-world** scenarios.

- However, it is never possible to include all of the various **bio-sensing** modalities in a **compact wearable** manner. Thus, this dissertation focused on two **real-world applications** to compare the performance of some widely-used sensor modalities.

- The hardware and software frameworks developed above are **modular, scalable,** and **robust** making them easily expandable to other **affective computing** applications.

PUBLICATIONS

Journals

Siddharth and Mohan M. Trivedi. "On Assessing Driver Awareness of Situational Criticalities: Multi-modal Bio-Sensing and Vision-Based Analysis, Evaluations, and Insights." *Brain Sciences* 10, no. 1, 2020.

Siddharth, Tzyy-Ping Jung, and Terrence J. Sejnowski. "Impact of Affective Multimedia Content on the Electroencephalogram and Facial Expressions." *Nature Scientific Reports* 9, no. 1, 2019.

Siddharth, Tzyy-Ping Jung, and Terrence J. Sejnowski. "Utilizing Deep Learning Towards Multi-modal Bio-sensing and Vision-based Affective Aomputing." *IEEE Transactions on Affective Computing,* 2019.

Siddharth, Aashish N. Patel, Tzyy-Ping Jung, and Terrence J. Sejnowski. "A Wearable Multi-modal Bio-sensing System Towards Real-world Applications." *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 4, pp. 1137-1147, 2018.

Conferences

Siddharth and Mohan M. Trivedi. "Attention Monitoring and Hazard Assessment with Bio-Sensing and Vision: Empirical Analysis Utilizing CNNs on the KITTI Dataset." In *2019 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1673-1678, 2019.

Julia Anna Adrian, **Siddharth**, Syed Zain Ali Baquar, Tzyy-Ping Jung, and Gedeon Deák. "Decision-Making in a Social Multi-Armed Bandit Task: Behavior, Electrophysiology and Pupillometry." *41st Annual Meeting of the Cognitive Science Society (CogSci)*, 2019.

Siddharth, Tzyy-Ping Jung, and Terrence J. Sejnowski. "Multi-modal Approach for Affective Computing." In *IEEE 40th International Engineering in Medicine and Biology Conference*, 2018.

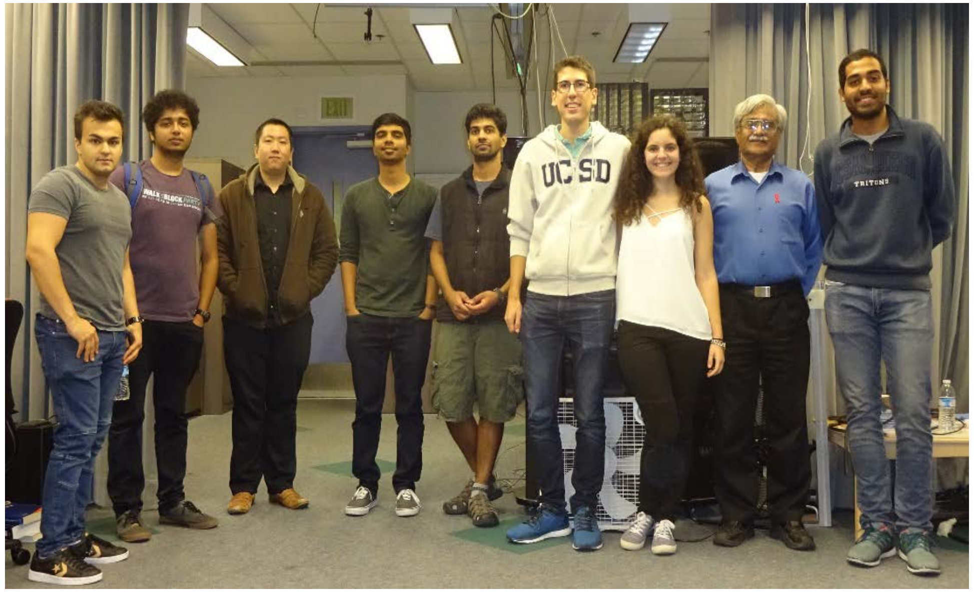
Siddharth, Aashish N. Patel, Tzyy-Ping Jung, and Terrence J. Sejnowski. "An Affordable Bio-sensing and Activity Tagging Platform for HCI Research." In *International Conference on Augmented Cognition*, pp. 399-409, 2017.

Siddharth, Akshay Rangesh, Eshed Ohn-Bar, and Mohan M. Trivedi. "Driver Hand Localization and Grasp Analysis: A Vision-based Real-time Approach." In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2545-2550, 2016.

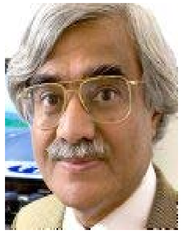
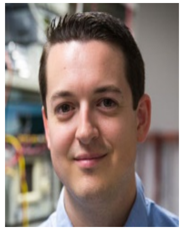
THANK YOU

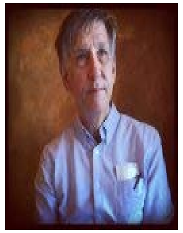
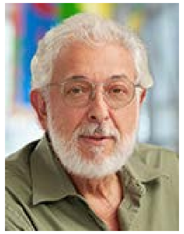
Swartz Center for Computational Neuroscience (SCCN)

Facebook Reality Labs (FRL)



Laboratory for Intelligent and Safe Automobiles (LISA)

NSF

ARL

Center for Wearable Sensors

FRONTIERS OF INNOVATION FISP SCHOLARS PROGRAM

CHANCELLOR'S RESEARCH CRES EXCELLENCE SCHOLARSHIPS

Kavli Institute for Brain & Mind