Exercise 2 Report: Physiological signals in affective computing

Rayan Armani*
rarmani@student.ethz.ch
ETH Zurich
Switzerland

Massimo Albarello*
malbarello@student.ethz.ch
ETH Zurich
Switzerland

1 INTRODUCTION

Knowledge of the emotional state of a user is key in affective computing, in order to create novel and improved user experiences. In this paper we look into the use of physiological signals collected from commercial sensors to determine the level of valence and arousal of a person. This report focuses on data processing and feature extraction leading up to training a random forest (RF) classifier to perform the emotion recognition task.

2 DATA PROCESSING AND FEATURE EXTRACTION

We used the ASCERTAIN Database [11], which relates data from several sensors (electroencephalogram (EEG), electrodermal activity (EDA), electrocardiogram (ECG) and facial landmark trajectories (EMO) signal) to emotion self-ratings of valence and arousal of participants, collected after watching video clips meant to trigger different emotional responses. The sensors used for data collection are off-the-shelf and prone to noise and artifacts, meaning data has to be preprocessed before extracting features relevant to emotion recognition.

2.1 ECG

2.1.1 Filtering: To limit the impact of noise in single lead ECG measurements, we took the Lead I ECG, which is the difference between left and right arm signals. We then retrieved the signal's power spectral density (PSD) using Welch's method as in [1], with a window window length of 15 × *sr* and the overlap of 10 ×*sr* , where *sr* is the signal sampling rate.

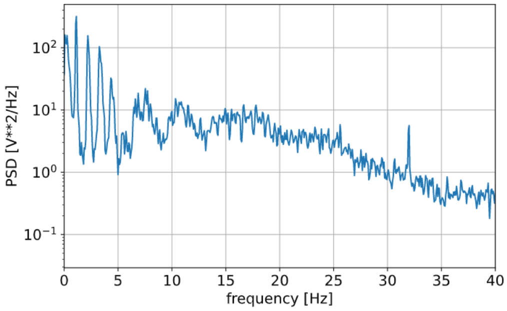

Figure 1: Lead I ECG Power Spectral Density plot [participant 1, clip 1]

Key ECG features are contained in the [0.05-35]Hz frequency band with QRS complex information in the [5-30] Hz band. However, the PSD in figure 1 shows both high and low frequency noise components that produce a wandering of the baseline of the signal. We implemented a bandpass filter in the [4-20]Hz range to balance information retention and noise reduction, as seen in figure 2.

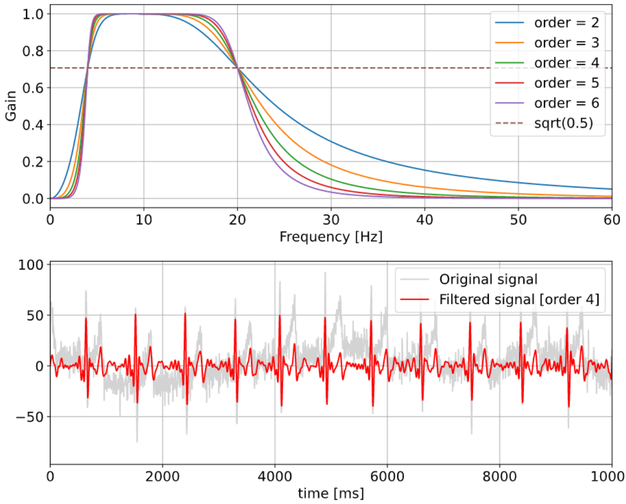

Figure 2: Frequency response of bandpass filter and filtered Lead I ECG [participant 1, clip 1]

However, as mentioned in [10], the very low frequency components of the PSD are relevant to emotion recognition so a low pass filter at 20Hz was used to extract features in the frequency domain.

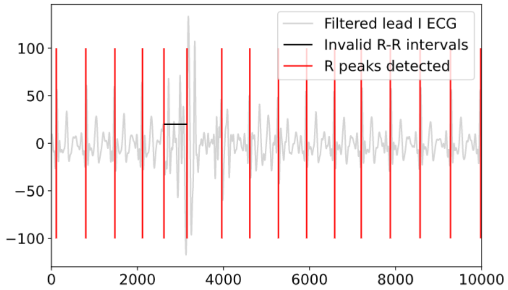

Figure 3: Artifact detection identifies invalid IBI interval [participant 7, clip 13]

2.1.2 Artifact detection: We implemented the algorithm detailed in [5] to detect artifacts in the Lead I ECG signals and flag invalid Inter-beat intervals (IBI) as shown in figure 3.

The percentage of recorded ECG data that contains artifacts (time of all flagged IBI) over the complete dataset (last 50 s for each clip) is around 6.56%. In cases where the entire signal is corrupted, the whole 50s were counted as artifacts.

To preserve the frequency spectrum of the ECG signal after artifact detection, we replaced artifacts with values given by linear interpolation between the closest valid R peaks. We then smoothed out remaining irregularities using adaptive thresholds around the IBI mean in a sliding window, replacing each outlier with the respective mean.

*Both authors contributed equally

Table 1: Features Extracted

Signal	Features
ECG	10 Low freq. ([0−2.4] Hz)PSDs; 4 very low freq.([0−0.04] Hz) PSDs; Statistical measurements over inter beat intervals, heart rate and heart rate variability
EMO	Vertical deformation of: Upper lip, Lower Lip, Left Lip corner and Right Lip corner; Horizontal deformation of: Left Lip corner and Right Lip corner; Deformation of the: Right Eyebrow, Left Eyebrow, Right Cheek, Left Cheek, Right Lid and Left Lid
EDA	Mean of: skin resistance (SR), SR first derivative, absolute values of SR first derivatives, negative values of SR first derivative; percentage of time with negative first derivative of SR; standard deviation of SR, # of local minima in the skin conductance (SC) signal, average rising time of the EDA signal; 4 PSD estimates in the [0-0.4] Hz band, standard deviation of SC; mean of: first derivative of SC, absolute values of SC first derivative, absolute values of SC second derivatives; # of local minima in the SR signal; 10 log PSD in the [0-2.4] Hz band; mean SCR peak amplitude, mean SCR rise time, mean SCR recovery time, area under the curve of phasic component, mean tonic signal
EEG	Statistical measurements for channels 1-8

Statistical measurements include: mean(μ), standard deviation (σ), skewness, kurtosis, % times the value is above μ + σ and below μ - σ

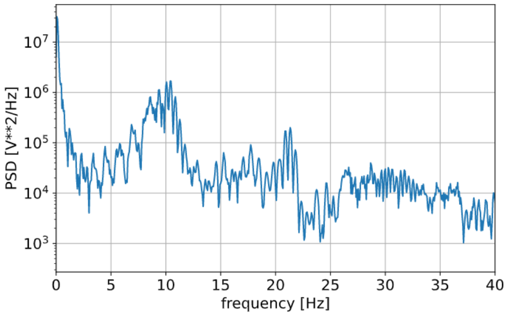

Figure 4: EDA Power Spectral Density plot [participant 2, clip1]

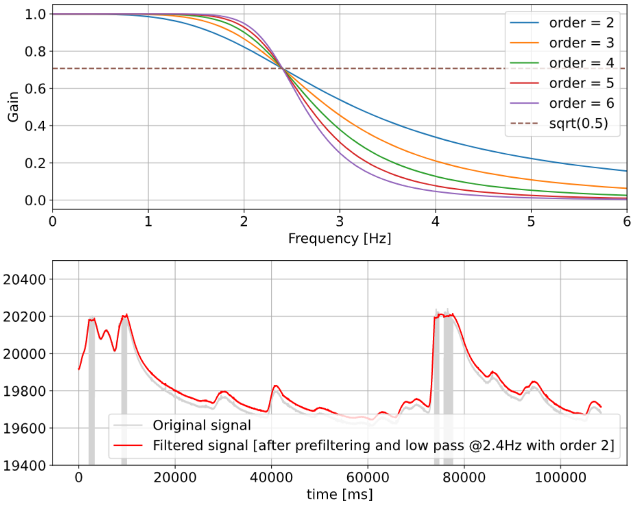

Figure 5: Filter frequency response and filtered EDA [participant 2, clip 1]

2.1.3 Feature extraction: In addition to features extracted in [11], we also extracted the following:

- Root mean squared IBI: as explained in [10], this is the primary time-domain measure used to estimate the changes reflected in HR Variability
- Number of adjacent IBI that differ by more than 50 ms: can characterize how 'dynamic' the emotional reaction is
- percentage of adjacent IBI that differ by more than 50 ms: complementary to the one above,it can show how much time the person experience changes in HR.
- HR max - HR min: a bigger range of HR might point to a change in emotions, whereas a narrower one could hint at a more stable emotional response.
- max HRV (absolute value): same logic as above.

2.2 EMO

Features were extracted from a subset of interest points and are summarized in table 1.

2.3 EDA

2.3.1 Filtering. We obtain the PSD of the EDA signal with the same method used for the ECG signal.

The power spectrum (Figure 4) shows that the most significant part of the signal is concentrated in the low frequencies which is expected since EDA is a slow response signal. We implement a low pass filter at 2.4Hz cutoff, meant to include the features most relevant to emotion recognition.

A simple low pass filter however does not eliminate the artifacts that can be noticed in some of the raw EDA recordings. We therefore added a prefiltering step: after applying a 25Hz low pass filter to the derivative, we threshold around the derivative and replace outliers by a random variable sampled from a truncated normal distribution mirroring the inherent noise so as to not alter spectral features. The values of the thresholds and low pass filter cutoff were determined empirically, to achieve results as shown in figure 5

2.3.2 Feature Extraction. As for the ECG, base features were extracted as in [11] and [1], and are listed in table 1. The EDA signal can be further decomposed into a tonic component, that represents the slow shifts in a person's base skin conductance levels, usually due to ambient factors; and a phasic component, which reflects more abrupt changes in skin conductance in response to emotions. We can extract additional features from the decomposed signal, using open source EDA processing libraries ([3], [7]) and insights from [12]:

- Mean amplitude of the skin conductance responses (SCR) peaks in the phasic signal (peak height - onset height): shows the amplitude of the emotional response
- Mean rise time of an SCR peak in the phasic signal: translates the speed of the emotional response
- Half recovery time of the SCR: describes how the emotional response tapers
- Area under the curve for the phasic component: can be a proxy of overall dermal activity due to emotional responses
- Mean of the tonic component: understanding and eventually cancelling out the base skin response

Table 2: Mean cross validation scores: all features | top 10

Cross val. scheme	F1score	Precision	Accuracy	Recall
Valence LOCO	0.43 | 0.45	0.50 | 0.56	0.55 | 0.59	0.57 | 0.59
Valence LOPO	0.45 | 0.53	0.53 | 0.56	0.53 | 0.57	0.51 | 0.57
Arousal LOCO	0.76 | 0.77	0.66 | 0.67	0.64 | 0.65	0.93 | 0.94
Arousal LOPO	0.45 | 0.52	0.56 | 0.57	0.55 | 0.55	0.50 | 0.57

Note: score for all features | score for top 10 features

2.4 EEG

The features extracted for EEG consist of statistical measurements across 8 EEG channels. They were further processed by imputing missing entries with the corresponding feature mean.

2.5 Valence and Arousal

Calculating the Pearson coefficient between the self reported valence and arousal ratings results in a value of -0.0177. Being very close to zero, it confirms that the two metrics are not strongly correlated and that the ratings can be used as separate dimensions to evaluate the emotional state of the participant.

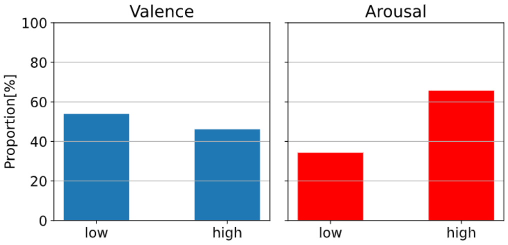

Figure 6: Class distribution of Valence and Arousal labels

We translate the emotion recognition task into a binary classification task by regrouping the valence and arousal labels into 'LOW' and 'HIGH' classes based on the middle of the respective range. The class distribution (figure 6) shows that while the valence dataset seems roughly balanced, the arousal dataset is imbalanced, with 'HIGH' arousal recordings being more represented than low arousal ones, 65.7% to 34.3%. This needs to be taken into account when training the classifier.

3 CLASSIFICATION

Using the open source library scikit-learn ([8]) we trained separate classifiers for valence and arousal by implementing both leave-one-clip-out (LOCO) and leave-one-participant-out (LOPO) cross validation. The models were evaluated using their accuracy, precision, recall, F1 scores (table 2); and tuned using a parameter grid search followed by manual adjustments. After a first training on all features, we extracted the 10 features with the highest importance (table 4) and re-iterated the cross validation schemes. While slight improvements were observed when using the reduced feature set, table 3 shows that the valence classifiers barely outperform the Zero Rule algorithm in accuracy, while the arousal classifiers fail to perform better. A deeper look into the confusion matrices in figure 7 reveals that the classifier results in both too many false positive and false negative predictions to be used in practice for emotion recognition.

Table 3: Accuracy comparisons

Cross val. scheme	ZeroR	Classifier Overall	Classifier VH/VL
Valence LOCO	0.54	0.59	0.58
Valence LOPO	0.54	0.57	0.52
Arousal LOCO	0.66	0.65	0.80
Arousal LOPO	0.66	0.55	0.53

Note: VH/VL regroups very high and very low self reported ratings

Table 4: Most significant features

Cross val. scheme	Best 10 features
Valence LOCO	mean, stdev and skewness of vertical deformation of Left Lip Corner, mean and stdev of Left Cheek, mean vertical deformation of Right Lip Corner, stdev of horizontal deformation of Right Lip Corner, mean of EEG channel 5, mean of the derivative of EEG channel 8, stdev of skin resistance
Valence LOPO	mean and stdev of vertical deformation of Left Lip Corner, stdev of Left Cheek, mean vertical deformation of right lip corner, mean HRV, kurtosis of IBI, mean of the derivative and proportion of negative derivative for EEG channel 8, proportion of values above μ + σ for EEG channel 1, mean log PSD of the EDA signal between [0.72,0.96Hz)
Arousal LOCO	mean derivative of skin resistance, average rise time and percentage of negative derivative of the EDA signal, kurtosis, proportion of values above μ + σ and below μ - σ for EEG Channel 1, mean HRV, mean deformation of right eyebrow, mean horizontal deformation of Left Lip corner, mean PSD of leadI ECG between [0.96,1.2Hz)
Arousal LOPO	mean and stdev of vertical deformation of Left Lip Corner, mean, stdev and skewness of vertical deformation of Right Lip Corner, mean of EEG channel 8 and proportion its derivative is negative, stdev of deformation of left cheek, stdev of Right eye lid deformation, mean log PSD of the EDA signal between [0.72,0.96Hz)

Note: score for all features | score for top 10 features

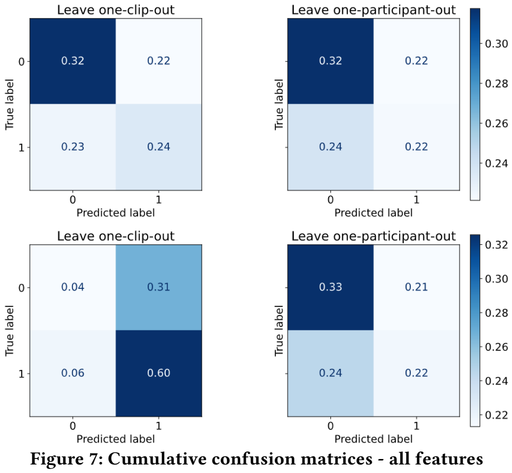

Figure 7: Cumulative confusion matrices - all features

4 DISCUSSION

4.1 Feature selection

While the exact top features are different across valence and arousal classifiers, table 4 shows that the features relevant to facial landmark trajectories such as statistical measurements of lips and cheek deformations generalise more. Implementing EMO measurements is simple but limited because in real-life settings users are facing a camera/webcam only for a limited amount of time. While some EEG features appear in the top 10, the measurement setup necessary for an 8 channel EEG is very sensitive to motion and noise and isn't practical enough for an "in the wild" application. The same holds for ECG and EDA sensors; indeed, even though these two would be the most practical ones for continuous tracking, the fact that they do not provide insightful features (in our classifier) make these sensors not suitable for emotion recognition in real life.

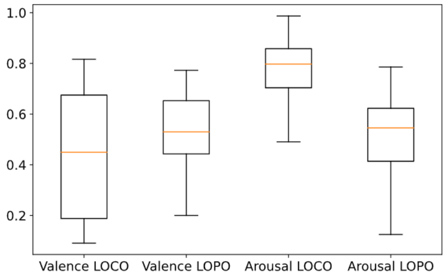

Figure 8: Spread of F1 scores across clips and participants

4.2 Performance and generalisation

The classifiers for valence and arousal perform differently in LOPO and LOCO cross validation. As show in figure 8, the valence classifier performs and generalises slightly better over participants than over clips, as the spread and standard deviation of F1 score is smaller in the LOPO scheme. The arousal classifier performs much better when trained over clips rather than people, and generalises better as well with a smaller spread. Intuitively, we expect our classifier to perform better in clips with extreme emotion ratings - very high [5, 6] and very low [0,1] arousal, very high [2,3] and very low[-3,-2] valence) - as we expect physiological signals to be more distinct. Results in table 3 however show that this is true only in the case of LOPO cross validation for arousal classification, where the accuracy of the classifier on the subset of very high or very low arousal ratings is higher than on the overall subset. In other cases, the accuracy is comparable or lower, indicating that our features or our classifier or both fail to correlate the intensity variations of physiological changes to the degree of emotions experienced.

4.3 Improving classification accuracy

Classification accuracy could be further improved with better feature engineering and/or selecting additional sensing modalities that encode more information about the emotional state of a participant. A chest worn IMU could be an interesting sensing modality because it can measure both breathing (movement of chest wall) and motion (fidgeting, laughing, etc.) . In fact, works such as [2] use a shirt mounted IMU used in emotional state classification in immersive experiences, and [4] classifies emotions from walking gait also measured with a chest IMU, hinting that this sensing modality could be investigated also in settings outside the lab.

Another modality to explore could be facial Electromyography (EMG) which measures the electrical activity in muscles and has been used in several other studies ([6], [9]. Wearing EMG electrodes however is not practical in real life settings, so this sensing modality would have limited applications. Instead, a more unobtrusive sensor such as a microphone could be used to improve emotion detection. A microphone could be used to measure breathing when in contact with the chest but also used to extract information from speech. Audio can be processed using natural signal processing to extract more sophisticated insights on the emotional status of the user.

REFERENCES

[1] Mojtaba Khomami Abadi, Ramanathan Subramanian, Seyed Mostafa Kia, Paolo Avesani, Ioannis Patras, and Nicu Sebe. 2015. DECAF: MEG-Based Multimodal Database for Decoding Affective Physiological Responses. *IEEE Transactions on Affective Computing* 6, 3 (2015), 209–222. https://doi.org/10.1109/TAFFC.2015.2392932

[2] Alberto Betella, Riccardo Zucca, Ryszard Cetnarski, Alberto Greco, Antonio Lanatà, Daniele Mazzei, Alessandro Tognetti, Xerxes D. Arsiwalla, Pedro Omedas, Danilo De Rossi, and Paul F. M. J. Verschure. 2014. Inference of human affective states from psychophysiological measurements extracted under ecologically valid conditions. *Frontiers in Neuroscience* 8 (2014), 286. https://doi.org/10.3389/fnins.2014.00286

[3] Alberto Greco, Gaetano Valenza, Antonio Lanata, Enzo Pasquale Scilingo, and Luca Citi. 2016. cvxEDA: A Convex Optimization Approach to Electrodermal Activity Processing. *IEEE Transactions on Biomedical Engineering* 63, 4 (2016), 797–804. https://doi.org/10.1109/TBME.2015.2474131

[4] Muhammad Arslan Hashmi, Qaiser Riaz, Muhammad Zeeshan, Muhammad Shahzad, and Muhammad Moazam Fraz. 2020. Motion Reveal Emotions: Identifying Emotions From Human Walk Using Chest Mounted Smartphone. *IEEE Sensors Journal* 20, 22 (2020), 13511–13522. https://doi.org/10.1109/JSEN.2020.3004399

[5] Karen Hovsepian, Mustafa al'Absi, Emre Ertin, Thomas Kamarck, Motohiro Nakajima, and Santosh Kumar. 2015. cStress. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*. ACM Press. https://doi.org/10.1145/2750858.2807526

[6] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. 2012. DEAP: A Database for Emotion Analysis ;Using Physiological Signals. *IEEE Transactions on Affective Computing* 3, 1 (2012), 18–31. https://doi.org/10.1109/T-AFFC.2011.15

[7] Dominique Makowski, Tam Pham, Zen J. Lau, Jan C. Brammer, François Lespinasse, Hung Pham, Christopher Schölzel, and S. H. Annabel Chen. 2021. NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior Research Methods* (02 Feb 2021). https://doi.org/10.3758/s13428-020-01516-y

[8] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[9] Wataru Sato, Koichi Murata, Yasuyuki Uraoka, Kazuaki Shibata, Sakiko Yoshikawa, and Masafumi Furuta. 2021. Emotional valence sensing using a wearable facial EMG device. *Scientific Reports* 11, 1 (March 2021). https://doi.org/10.1038/s41598-021-85163-z

[10] Fred Shaffer and J. P. Ginsberg. 2017. An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health* 5 (2017), 258. https://doi.org/10.3389/fpubh.2017.00258

[11] Ramanathan Subramanian, Julia Wache, Mojtaba Khomami Abadi, Radu L. Vieriu, Stefan Winkler, and Nicu Sebe. 2018. ASCERTAIN: Emotion and Personality Recognition Using Commercial Sensors. *IEEE Transactions on Affective Computing* 9, 2 (2018), 147–160. https://doi.org/10.1109/TAFFC.2016.2625250

[12] Rafael Wampfler, Severin Klingler, Barbara Solenthaler, Victor Schinazi, and Markus Gross. 2019-07. Affective State Prediction in a Mobile Setting using Wearable Biometric Sensors and Stylus. In *Proceedings of the 12th International Conference on Educational Data Mining, EDM 2019, Montréal, Canada, July 2-5, 2019. International Educational Data Mining Society (IEDMS) 2019*, Michel C. Desmarais, Collin F. Lynch, Agathe Merceron, and Roger Nkambou (Eds.). Université du Québec; Polytechnique Montréal, Montréal, 198 – 207. https://doi.org/10.3929/ethz-b-000393912 12th International Conference on Educational Data Mining (EDM 2019); Conference Location: Montreal, Canada; Conference Date: July 2-5, 2019; Conference lecture held on July 5, 2019.