ÉCOLE NORMALE SUPÉRIEURE/PSL,
UNIVERSITÉ DE PARIS, EHESS

MASTER IN COGNITIVE SCIENCES
2020-2021

M2 THESIS

Predictive modeling of arousal:

an analysis of a public EEG and EDA database

Author:
Tomás Ariel D'AMELIO

Tutor:
Dr. Denis-Alexander ENGEMANN

Laboratory:
Parietal
INRIA Saclay

Abstract

Arousal is implicated in different processes such as vigilance, alertness, attention, and affective states. Several studies have examined the correlates of this phenomenon, both at the central and peripheral nervous system levels. However, the configuration of the coupling dynamics between neural activity and behavioural measures related to this concept remains to be unravelled. One approach to address this issue is to implement predictive modelling of subjective arousal from electroencephalography data. However, it is possible that this bypassing of peripheral information to explain self-reported subjective arousal is missing out on available information to further understand this process. Thus, a two-step approach was proposed in the present thesis. Firstly it was assessed whether it is possible to decode electrodermal activity from EEG. Then, it has been evaluated if the predicted electrodermal activity can work as a proxy measure and have additive effects in the subjective arousal prediction task. This way, the first objective was achieved: it was possible to implement a regression pipeline that effectively predicts the electrodermal activity at the event-level with the data obtained from EEG recording. However, no additive effects were found in the explanation of affective arousal by the predicted electrodermal activity signal. Importantly, a programmatically-based methodology for predictive arousal modelling was established in the present thesis. This would allow extending this to other peripheral measures and other databases to achieve a better understanding of the arousal phenomenon.

Keywords: Arousal decoding, Predictive modelling, EEG, EDA

Declaration of Novelty

This thesis establishes a novel methodology for understanding the dynamics between central and peripheral nervous system signals, and their corresponding behavioural measures. It is proposed that this methodology could be of particular relevance for the understanding of arousal, considering the link between the central and autonomic sources related to this process.

Possibly the most important aspect of this work is that it has indeed been possible to predict the electrodermal activity signal from the electroencephalography data of the participants, a task that has not been previously addressed in the literature. Furthermore, considering linear models were implemented, it has been possible to observe patterns of activity at the sensor level in terms of the main components of the predictions.

At the same time, another innovative aspect of the thesis has been to implement

statistical models that consider possible additive components between actual and predicted peripheral signals, although this has not been found at least in relation to the variance of electrodermal activity.

Declaration of contribution

The whole work that is comprised in this thesis was done by me (TD) and Dr. Denis-Alexander Engemann (DAE). The contribution made by each of them is detailed below.

- Definition of the theoretical question: TD; DAE
- Bibliographic research: TD
- Definition of general approach: TD; DAE
- Definition of the methods: TD; DAE
- Preprocessing EEG, EDA and EMG data: TD
- Analysis: TD
- Interpretation of the data: TD; DAE
- Thesis writing: TD
- Thesis correction: DAE

Acknowledgements

First of all, I would like to thank my supervisor, Denis Engemann. Not only for helping me to confirm my interest in science in the broadest and truest sense of the word, but also for his concern for my well-being. This journey would not have been the same without you.

I would also like to thank my *alma mater*, the University of Buenos Aires. As a public, free and high-quality university that has undoubtedly fostered my academic growth in every sense. My commitment will always continue to ratify the scientific character of psychology in our country.

I would like to thank the members of Neurotransmitiendo. Those who guide me day by day to be a better professional.

Finally, I would like to thank my family and loved ones for the unconditional support they always give me.

Gracias totales!

Contents

1 Introduction

1.1 Literature review

1.1.1 A neurophysiological-grounded approach for emotion research

A vast number of models is used for the conceptualization and measurement of emotions. One issue that several models of emotion have in common relies on comprehending affective states as adaptive responses taken by organisms towards a given context. In this way, some authors focus on the importance given to the aspect of emotion expression as an intrinsic part of what emotion consists of (Hess and Thibault, 2009). In this way, the first models of basic emotions emerged, which proposed the existence of categories of emotions (i.e. happiness, sadness, fear, disgust, anger and surprise; Ekman and Friesen, 1971). Along with this idea, one of the key propositions of this model relies on distinct neurological patterns specific to these basic emotions (Ekman, 1992). However, mapping independent neural patterns for the different basic emotions in humans has been a major challenge for the adherents of this model, generally arriving to inconsistent findings (Posner et al., 2005). As described by Kragel and LaBar (2016), no conclusive results have been found that allow a one-to-one mapping of the different basic emotions from univariate analyses. Instead, performing multivariate analyses (e.g. multivoxel pattern analysis) would allow differentiating between discrete emotions (Kragel and LaBar, 2016), although this neglects the principle that basic emotions have independent discrete patterns. This opens the possibility towards other models of affective states as more explanatory of this phenomenon.

A such, the circumplex model of emotion (Russell, 1980) is a dimensional approach that aims to solve these difficulties. By resuming the ideas outlined by Wundt (1912) and Schlosberg (1952), this model posits that all emotional states emerge from the interaction between two principal components, one being valence (between pleasure and displeasure) and the other one being arousal (Posner et al., 2005). At the neural level, this means that affective states would arise from two overlapping neurophysiological systems (Wilson-Mendenhall et al., 2013). In

exploring this issue, Colibazzi et al. (2010) have indeed found distinct patterns of activation of neural structures that are consistent with the dimensional model of affective states. This way, while the dorsal and mesolimbic cortex were observed to be related to valence coding, the amygodalo-hippocampal-parahippocampal system among other subcortical structures (e.g. thalamus, globus pallidus and caudate) were related to arousal (Colibazzi et al., 2010).

At the same time, strong evidence for this two-dimensional model is found in self-report data, mainly through factor analysis and multidimensional scaling of reported affective states (Barrett and Fossum, 2001; Posner et al., 2005). Thus, the experience of a discrete emotion would be nothing more than the interpretation of the perceived arousal and valence at a given moment (Posner et al., 2005). Therefore, we could differentiate between two components of affective states: a *core affect*, as that basic neurophysiological state that the organism experiences, and *conceptual knowledge*, as something that is learnt and consequently applied to the acknowledgement of the core affect (Barrett, 2006). The core affect is compatible with the dimensional model of affective states just described, and would be useful to understand the primary components of the affective states of organisms.

In the same way, different authors consider a neurophysiological approach would aid in the understanding of the sources involved in affective processes. For example, LeDoux (2012) also bases itself upon basic circuits present in humans as well as non-human animals but discards the concept of emotion to refer instead to *survival circuits* (LeDoux, 2012). These would be the recognisable circuits underlying what is commonly understood by emotion across species. Survival circuits allow phenomena like emotion, motivation and arousal, for example, to be seen as parts of a complex process that comes into play when challenges or opportunities appear (LeDoux, 2012). In order to respond to a relevant environmental stimulus, it is important for an organism to have systems that allow it to recognize it, prepare itself to either approach or avoid it, and consolidate whether it is important for survival and therefore not forgotten. It is at this point when arousal acquires particular relevance.

1.1.2 Arousal state

Arousal is a conceptually complex term. It not only refers to an affective dimension but it could also be defined as a global activation state and related to wakefulness. There are different structures (e.g: the reticular formation in the midbrain tegmentum) that participate in the sleep-wake cycle regulation and in the general level of activation, as well as in attention, muscle tone and other related mechanisms (Halász et al., 2004). Additionally, arousal is known to be one of the principal dimensions of consciousness. In fact, the relationship between arousal and awareness can be seen in unconscious states like anaesthesia and non-REM sleep. Moreover, under pathological conditions such as narcolepsy, not only various aspects of the sleep-wake cycle are prone to changes, but autonomic and affective arousal are also affected (Schiappa et al., 2018). Thus, it can be observed that there is a relationship between the mechanisms that regulate wakefulness, attention and arousal as a physiological state.

The relationship between affective arousal and autonomic and wakeful arousal aspects was developed by the pioneer theorists of dimensional models of affective states. In this way, Russell (1980) uses arousal as a synonym of activation and opposes it to sleep as the other extreme of the affective dimension. Similarly, the arousal dimension within one of the most cited affective self-reports, the self-assessment manikin (SAM; Bradley and Lang, 1994), is presented as the opposition between excited (i.e. stimulated, frenzied, jittery, wide-awake or aroused) and calm (i.e. relaxed, sluggish, dull, sleepy or unaroused). According to Russell and Barrett (1999) and Storbeck and Clore (2008), it would not be possible to separate affective and autonomic arousal, considering that the sympathetic and endocrine system activation is inherent to this emotional dimension. This raises the doubt on the relationship between the different types of arousal, and the possibility of taking advantage of possible common neural mechanisms between types of arousal for a better prediction of subjective arousal.

1.1.3 Arousal and stress

One of the concepts most closely related to the dimension of affective arousal is stress. Stress is an adaptive response to cope with relevant stimuli in their context. It prepares the organism to deal with the challenges presented by the environment, as well as triggering metabolic processes to having energy resources available to carry out such action. For example, in case a burglar unexpectedly breaks into the reader's house right now, this will inevitably generate a stress response that allows mobilizing resources in function of a certain way of coping with such a situation (e.g. running in the direction of a bedroom to lock oneself in and call the police). The mere thought of the possibility that this could happen to you would be enough to trigger such a stress response (Cisler and Koster, 2010). Stress responses are not only present in the case of aversive situations. We can also experience stress in response to appetitive stimuli (e.g. a date with a person one has been wanting to go out with for a long time), which would elicit similar hormonal responses (Merali et al., 1998). Thus, this activation of the organism that we understand as arousal makes sense as a particular dimension within the self-report of affective states since it explains a part of the emotional response that in principle is not explained by its valence(Yik et al., 1999).

At the physiological level it is possible to delimit which is the information pathway since this stressor event is processed until a certain response is consequently executed. In principle, sensory information reaches the brain mainly through the ascending pathways of the brainstem. One of the structures that primarily processes stressor stimuli, and which is most closely related to the response to salient stimuli, is the amygdala (Kukolja et al., 2008). The amygdala is related to the evaluation of possible presence of relevant stimuli in the environment, reacting to stimuli with social-emotional significance. Several studies have reported the role of different amygdalar nuclei in the onset of the stress response, whose connections with the hypothalamus trigger processes of the well-known hypothalamic-pituitary-adrenal axis (LeDoux et al., 1988; see Figure **??**). This axis consists of hormonal communications that begin with the production of corticotropin releasing hormone in the hypothalamic paraventricular nucleus, and ultimately enable increased cortisol production in the adrenal cortex . This glucocorticoid promotes glucose formation,

as well as intervenes to decrease glucose utilization in most cells, except for organs such as the brain and heart. This prioritizes the activity of these organs at the expense of other bodily functions.

Another product of the stress response relies on the release of catecholamines (e.g. norepinephrine; (Fuller et al., 2011)), related to the activation of the sympathetic adrenomedullar system axis. These processes increase arousal in the organism, as well as attention and vigilance, and lead to the known autonomic changes related to the stress response. For example, an increase of the heart rate can be registered, which is a consequence of the direct inhibition of the activity of the sympathetic nervous system by inhibition of the vagus nerve via GABAergic neurons (Cool and Zappetti, 2019). But it is also possible to enumerate other autonomic responses such as changes in muscle tone, as well as changes in sweat gland activity. The postganglionic sudomotor neurons have the particularity of being cholinergic, unlike the rest of the postganglionic sympathetic neurons which are noradrenergic (Ziemssen and Siepmann, 2019). The activity of these sweat glands was of great importance in psychophysiology since it was one of the first measures that could be obtained systematically through the so-called galvanometers. These devices that allow tracking electrophysiological changes resulting from sweating have been refined over time. Nowadays the measurement of skin conductance is one of the most important correlates of the stress response and arousal as an affective dimension (Boucsein, 2012).

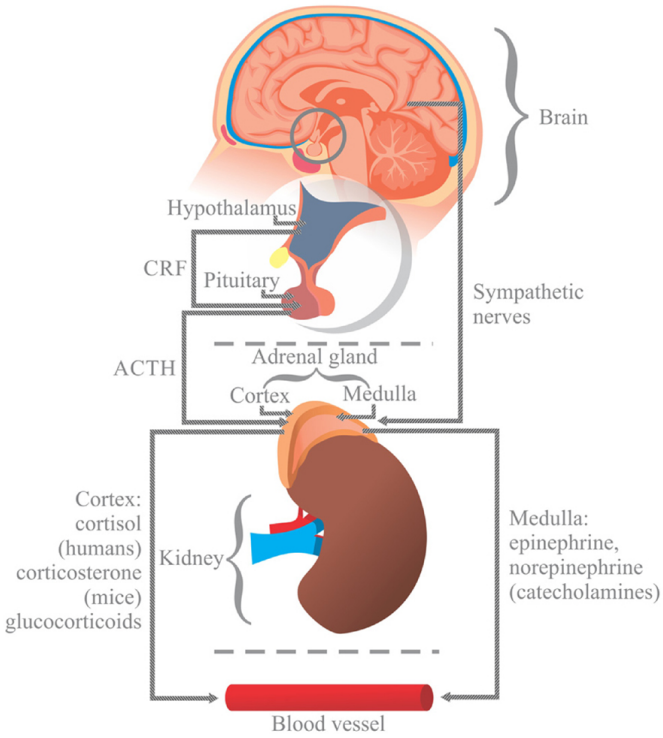

Figure 1. Activation of the hypothalamic-pituitary-adrenal and the sympathetic adrenomedullar axes. When the stress response is activated, the hypothalamus secretes the corticotropin releasing hormone (CRH), which activates the pituitary gland to produce another hormone called adenocorticotrophin (ACTH). This last hormone activates the adrenal gland which secretes cortisol, epinephrine and norepinephrine. This last two component secreted are of particular relevance because they are related to elicitation of peripheral response (e.g. changes in blood pressure, heart rate and sweating). Reprinted from Campos-Rodrıguez, R., Godınez-Victoria, M., Abarca-Rojano, E., Pacheco-Yepez, J., Reyna-Garfias, H., Barbosa-Cabrera, R. E., & Drago-Serrano, M. E. (2013). Stress modulates intestinal secretory immunoglobulin a. *Frontiers in integrative neuroscience*, *7*, 86, p. 3.

1.1.4 Sources of EDA

Electrodermal activity (EDA) is a general term which indicates variations in the electrical properties of the skin due to the activity of the eccrine sweat glands (Boucsein, 2012). These glands are directly innervated by the sympathetic branch of the autonomic nervous system and, more specifically, by the sudomotor nerve.

Changes in the skin conductance of particular sites of the human body, such as the fingers and palms, where the concentration of eccrine glands is higher than in other parts of the body, can be related to a person's psychophysiological state, as well as to their interaction with exogenous events (Greco, Valenza, Lanata, et al., 2015). The pathways involved in the central control of EDA are numerous and complex (Schell and Filion, 2007). Thus, different sources can be traced for the generation of this activity (see Figure 2).

In the first instance, skin conductance responses (SCRs) have been observed as a correlate of the evaluation of the relevance of a stimulus in a given context and as a correlate of the participants' own emotional experience. This activity has been found to be related to activation of the amygdala, hippocampus and anterior cingulate cortex(Boucsein, 2012). Another source of electrodermal activity which has common structures with the aforementioned source is the so-called 'thermoregulatory sweating' source, which correlates with hypothalamic activity. These two processes correspond to a first 'limbic-hypothalamic' source, as systematised by Boucsein (2012). It makes sense to think of the emotional and thermoregulatory sources as having similar sources if we consider that both processes involve an adaptation of the organism to an increase in its body temperature due to higher metabolic consumption during states of arousal.

In addition, it is possible to account for a second source of electrodermal activity related to the preparation of movements. The structures that would constitute this source are the pre-motor cortex and the basal ganglia. Finally, a third source of EDA is related to reticular activation(Boucsein, 2012).

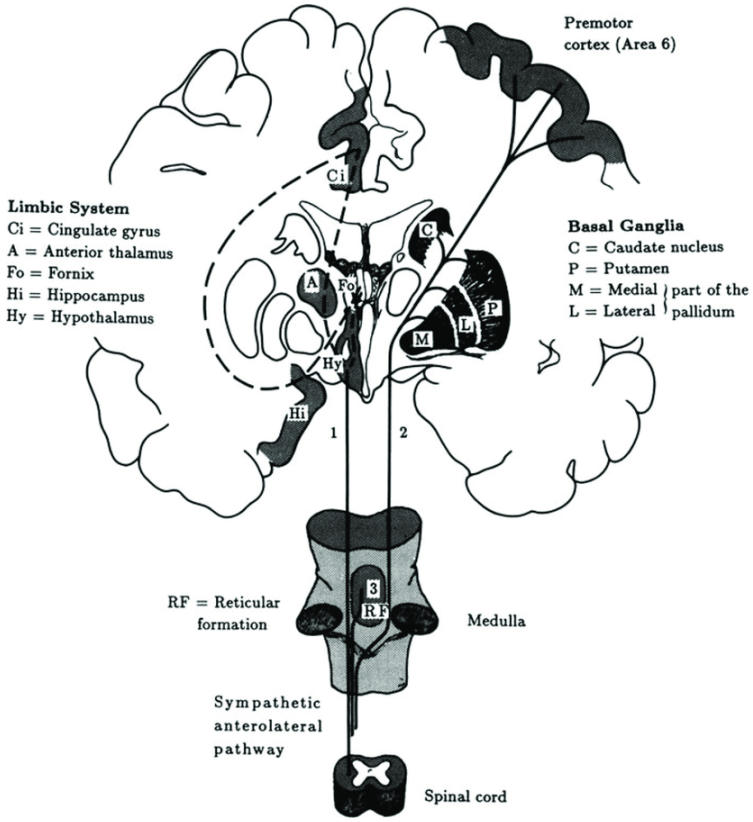

Figure 2. Central nervous system sources of EDA. 1: ipsilateral control of affective and thermoregulatory processes (cingulate gyrus, anterior thalamus, fornix, hippocampus, hypothalamus). 2: contralateral control of movement planification processes (premotor cortex and basal ganglia). 3: reticular formation influence. Reprinted from Boucsein, W. (2012). *Electrodermal activity*. Springer Science & Business Media, p. 40.

According to Sequeira and Roy (1993), there would be different levels of lateralization of electrodermal activity depending on the source of the signal. A first level of control would be associated with contralateral influences from the cortex and basal ganglia, whereas a second level of EDA would be involved with ipsilateral influences from the 'limbic-hypothalamic' source (Roy et al., 1993). One of the pioneer studies on this phenomena was conducted by Mangina and Beuzeron-Mangina (1996), reporting that the electrical stimulation of the left hemisphere's limbic structures produced greater relative activation of the left hand's EDA, whereas the analogue stimulation of the left hemisphere produced greater relative activation

of the right hand. Hence, it was hypothesized that the human generation of bilateral EDA in response to stimulation of amygdala and hippocampus structures was under an ipsilateral control (Mangina and Beuzeron-Mangina, 1996; Roy et al., 1993).

As we can notice, EDA is intimately related to arousal. Both in physiological terms, because of the structures whose stimulation produces changes in skin conductance; as well as functionally, considering that reported changes in arousal lead to differences in EDA. It is therefore hypothesized that a better understanding of signal generating mechanisms of EDA through predictive modelling could improve the predictive modelling of self-reported arousal in each participant.

1.1.5 Predictive modelling of arousal

Arousal prediction has become common within engineering and computer science literature in recent years. In these studies, EDA or EEG, among other signals, are used to recognize participants' self-reported affective states during an emotion elicitation task. A pioneering study has been carried out by Kim and André (2008), who aimed to classify the participants' affective states based on multiple peripheral measures, using both subject-dependent (i.e. train and test on the same subject) and subject-independent (i.e. train and test disregarding the subject) event-level models. As expected, better results were found in the intra-subject trained models. Furthermore, and consistent with prior literature, a better performance was achieved for arousal classification compared to valence. Similarly, Greco, Valenza, Citi, et al. (2016) also intended to classify arousal and valence in response to auditory emotional stimuli, but using only the EDA of participants as input for the models to be trained. Again, they found better results for arousal recognition. This is congruent with the idea that the estimation of arousal usually corresponds to sympathetic nervous system discharge, while the prediction of valence would emerge from the combination of multiple physiological measures (Kim and André, 2008).

The standard methodology in affective decoding research consists of supervised statistical learning models. These models are characterized by having each observation of its predictors $x_1, ..., x_n$ associated with a response measurement y_i. Thus,

the supervised models in machine learning imply fitting a model in order to find a function that allows representing the interaction between the different variables x_1 to predict a response y_i (James et al., 2013). In this way, as a first step of the supervised models for arousal decoding, participants' physiological activity averages are calculated in a feature engineering procedure. Then, a supervised model is trained, usually across-subjects. This involves estimating f, from parametric (i.e. establishing assumptions about f functional form) or non-parametric (i.e. without assumptions about f functional form) mappings of predictor variables, without considering that the trials come from different subjects. Finally, the prediction of participants' reported affective state is computed.

The implementation of event-level models predominates in literature, as it is difficult to achieve a sufficient sample size to be able to train robust subject-level models (e.g. prediction of emotional disorder risk scores). However, it is known that by using the subjects' mean as a representative value, information that could explain the intra-subject variability of the emotional phenomenon is lost (Fisher et al., 2018; Smith and Little, 2018). Thus, by using inter-subject models, it produces missing opportunities from available data in the aforementioned decoding task.

Consequently, in recent years there have been attempts to take advantage of individual subject signals to complement event-level models in affective states decoding tasks. In this sense, Arevalillo-Herráez et al. (2019) used an approach which complemented intra- and inter-subject information. The first finding of that study has been that, when training event-level models during emotional tasks, the contribution of the variable associated with the subject to the EEG signal is even greater than the emotional component itself. In this way, they noticed that using event-level models without the corresponding standardization would lead to classification that would not achieve its purpose.

However, the authors also mention that using subject-level models entails problems, such as the previously mentioned low sample size that usually have the training sets. Therefore, their proposal was to implement non-linear transformation to the features vectors, an alternative to the standardization from the linear transformations of the feature vectors that are usually done (e.g. z-score standardization). As a result of this combination of information, they found an improvement in the performance

of these models compared to both across-subject and intra-subject models. However, even considering only the experiments carried out on the databases with the highest performance, its achieved accuracy was not greater than 0.62 for arousal models with a binary codification of the response. Thus, the question of how to integrate inter and intra-subject information to improve performance in arousal decoding tasks remains open in the literature.

However, it is possible that one of the main problems of arousal decoding models is the way in which subjective arousal is measured. Considering that arousal is a dynamic and continuous concept, establishing a single measure of arousal at the end of each stimulus presentation possibly interferes with the possibilities of extracting information from the physiological signals Hofmann et al. (2020). Thus, a dynamic relationship between subjective arousal and EEG activity was evaluated using participants' continuous self-report as the target of the model. The statistical learning approach used involved both decoding high and low arousal states from Common Spatial Patterns (Ramoser et al., 2000) and Long Short-Term Memory recurrent neural networks (Hochreiter and Schmidhuber, 1997), as well as predicting *continuous* self-reports with Source Power Commodulation (Dähne et al., 2014) SPoC is a method for generating a reduced set of power features (from M/EEG neural oscillations) that are correlated with a target using supervised spatial filters. In this case, the target was the self-report of subjective arousal in an immersive VR experience. The SPoC method has been used previously for a variety of regression tasks. Among them, is possible to mention the prediction of brain age made by (Sabbagh et al., 2020). In this same work, predictions at the event-level were made for EMG decoding, a task that had also been previously performed by (Meinel et al., 2016). However, other methods also exist in the literature to statistically act on physiological signals. Among these, the Riemannian models, which seem to be better able to handle electromagnetic field spread, have now gained prominence(Sabbagh et al., 2020).

While self-reporting of subjective arousal is an optimal way to measure this affective dimension, the case of Hofmann et al. (2020) methodology is an isolated case in literature. In most databases containing arousal measures, this is reported only at the end of each trial. Is it possible to use these novel statistical learning

methods for arousal prediction from EEG in a way that takes advantage of all available information?

One possibility would be to learn to predict a widely available outcome and to exploit the correlation with the outcome of interest. This approach is known as *proxy measure*, which is nothing more than an approximation of signals from a given dataset containing both information from the predicted measure and from the selected feature set. In our case, a measure that could work as a proxy measure is the electrodermal activity, considering both its relationship with central generating sources (i.e. EEG) and arousal (target of our model), which has been previously used in the literature under tasks of prediction of brain age (Engemann, Kozynets, et al., 2020), but also of other cognitive variables such as neuroticism or fluid intelligence (Dadi et al., 2020). Thus, is it possible that this type of approach could work for a subjective arousal approximation from physiological data?

1.2 Research goals and hypotheses

In the present project, it is assumed that the information provided by EEG and EDA signals are closely related to the generation of arousal. However, it is also proposed that both signals would provide different information to describe the subjective arousal of each subject, by complementing peripheral information (i.e. EDA) with data related to the central nervous system (i.e. EEG).

Based on the two-step procedure it is intended to test these hypotheses. Thus, in the first place, it would be assessed whether it is possible to decode EDA from EEG. As EDA prediction from EEG is a task that has not been previously addressed in the literature, it is planned to first perform EMG prediction from EEG (previously performed in Sabbagh et al., 2020) to fine-tune the prediction pipeline prior to tackling the innovative EDA decoding task. Finding this result would imply that it is indeed possible to trace the generating sources of EDA through EEG (controlling by error sources as eye blinks).

On the other hand, it would be evaluated if the predicted EDA has additive value to subjective arousal prediction. If we find that, this would imply that the predicted EDA would have information that is not redundant for the arousal decoding task

compared to the real EDA.

2 Methods

We propose a methodology that aims to exploit intra and inter-subject information to:

A. Disentangle the relationship between the EDA response and EEG predictors.

B. Maximize the performance of self-reported arousal decoding models.

As opposed to the classical direct method of subjective arousal prediction from EDA or EEG, we propose an indirect two-step approach (see Figure 3). First, EDA decoding from EEG would be performed, with the idea of boosting high-resolution signals in each subject. Thus, we would train models with continuous inputs (i.e. EDA) and outputs (i.e. EEG) in order to extract as much information as possible from each subject to learn different subject-level function approximations. As an output of this first step, we would represent arousal with a predicted EDA version. This predicted signal would indirectly portray the coupling of autonomic and cerebral arousal, as it is the result of the sum of the EEG features weighted by the different coefficients generated after the fitting process with the EDA output data. This would allow making use of the richness of these signals in a unique representation at the subject level. In this way, the predicted arousal would contain information that is not included in the original EDA data. Consequently, in a second step, we would predict self-reported arousal from the predicted EDA, constituting the second part of the proposed statistical learning approach.

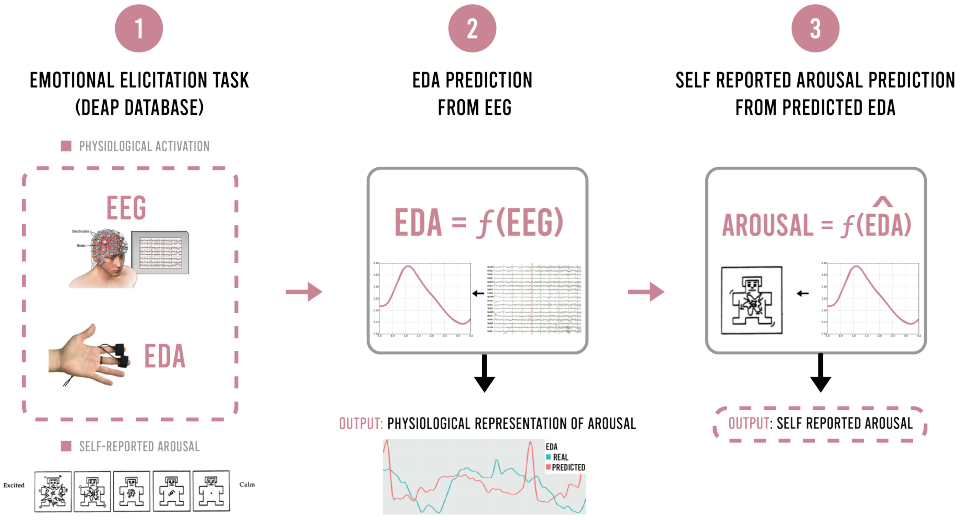

Figure 3. Proposed two-step approach for subjective arousal decoding.

2.1 Affective database

In this thesis, data provided by the Database for Emotion Analysis using Physiological Signals (DEAP; Koelstra et al., 2011) has been analyzed. To the time of this writing, this is one of the largest databases to study affective dimensions in a healthy population. The data were downloaded directly from the DEAP dataset website (http://www.eecs.qmul.ac.uk/mmv/datasets/deap/download.html), after completing and submitting the End User License Agreeement (EULA).

2.1.1 Participants

32 participants (16 women, mean age 26.9 years) were recruited for data collection (Koelstra et al., 2011). All of them signed a consent form prior to their participation.

2.1.2 Data acquisition

2.1.2.1 Physiological data acquisition

DEAP is a multimodal dataset containing different physiological signals (see Figure 4). The Biosemi ActiveTwo system was used to record:

- EEG (32 channels)
- EDA
- Horizontal and vertical EOG
- Zygomaticus major and trapezius EMG
- Respiration patterns
- Blood pressure (through plethysmograph)
- Skin temperature

All channels were recorded at 512 Hz. These recordings were obtained in two laboratories located at the University of Twente and the University of Geneva. Lighting was controlled, considering the effects that this could have on both the emotional experience and on the physiological signals of participants (Boucsein, 2012; LeGates et al., 2014). In addition, facial recordings of 22 participants were obtained while they were undergoing the experiment.

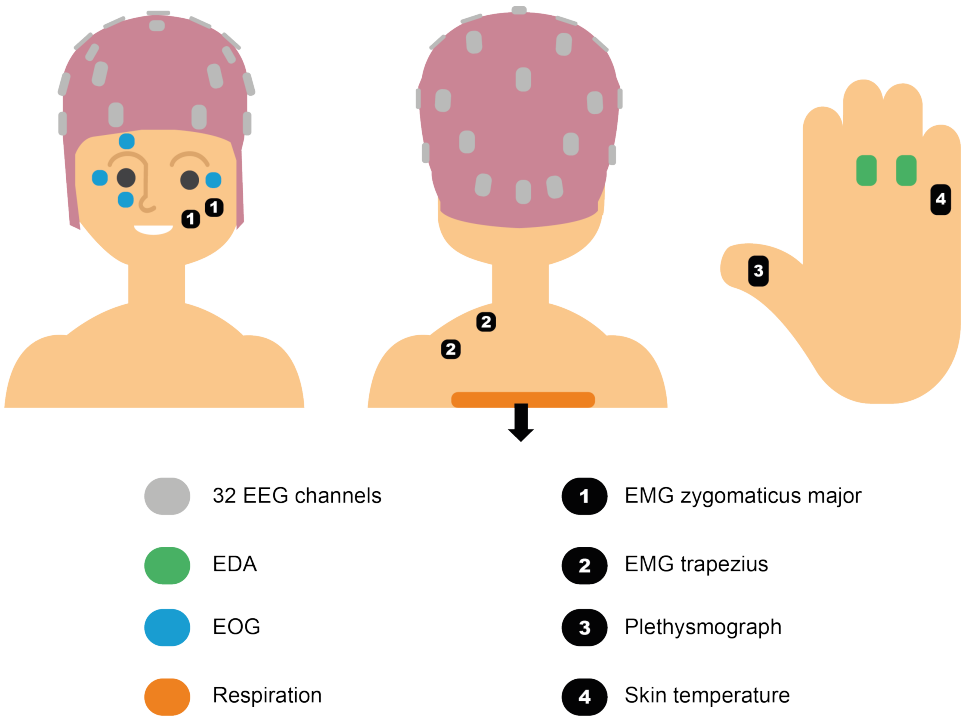

Figure 4. Location of physiological sensors in DEAP database. Adapted from Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., & Patras, I. (2011). Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, *3*(1), 18–31, p. 22.

2.1.2.2 Self-report data acquisition

Data collection for the DEAP database began with a practice trial to get participants familiarised with the task, followed by 2 minutes of baseline recording during which subjects were asked to relax (see Figure 5). Then, the task itself consisted of the presentation of the 40 one-minute emotional musical stimuli (with a break in the middle of the procedure), each followed by participants' self-report of the three dimensions of the Self Assessment Manikin: valence, arousal and dominance (Bradley and Lang, 1994). All scales were presented on a dimensional scale from 1 to 9. In this study, we focused on the arousal self-reports.

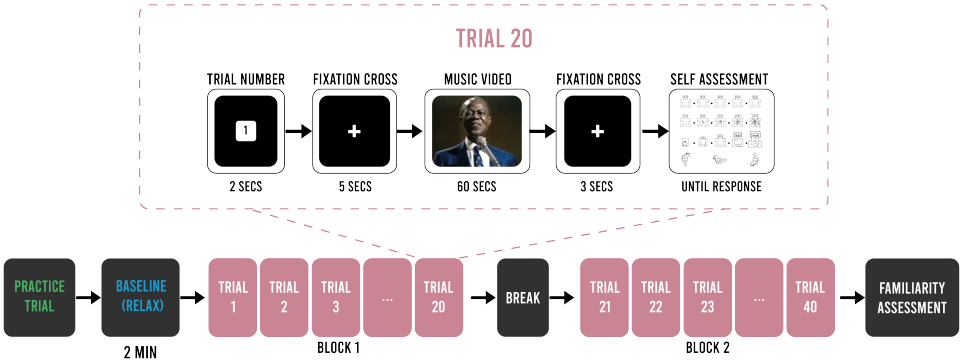

Figure 5. Representation of the procedure used in DEAP database (Koelstra et al., 2011).

In addition, a fourth scale was presented asking for the participant's liking (or disliking) of the video (i.e. a thumb up, a thumb to the side, or a thumb down) after each stimulus. Finally, at the end of the experiment, subjective data were obtained from participant's familiarity with each song presented.

2.2 Data processing

2.2.1 Software

Small-scale processing (e.g. data visualization) was performed locally on a Dell Inspiron 7386 computer (Intel Core i7, 16.0 GB RAM) with A Windows operating system. However, for more computationally demanding processing (e.g. EEG preprocessing or training machine learning models) a remote Linux server (72 cores, 376 GB RAM) was used. The connection to this server was made via Windows Subsystem for Linux (WSL) through a Secure Shell (SSH) protocol.

Data analysis was conducted using Python and R. All the scripts produced in this work can be accessed through a shared Github repository (see).

For EEG processing MNE v.0.23.0 (Gramfort et al., 2013) has been used. MNE (https://mne.tools) is an open-source Python software for human neurophysiological data processing, which is useful not only for exploring and visualizing physiological signals, but also for implementing statistical learning models from it.

On the other hand, Scikit-Learn library (Pedregosa et al., 2011) was used for machine learning analysis. Particularly for the implementation of predictive modelling from EEG we used Coffeine library (Sabbagh et al., 2020). General-purpose Python libraries such as NumPy (Harris et al., 2020), pandas (pandas development team, 2020) and Matplotlib (Hunter, 2007) were also used.

Finally, for data visualization and implementation of linear mixed models, ggplot2 (Hadley, 2016) and lme4 (Bates et al., 2007) R packages were used.

To guarantee transparency in data processing, code used for data analysis can be found on Code Availability section (see Section 5).

2.2.2 Data pre-processing

Although it was possible to work directly with DEAP pre-processed data in Python format (provided as a variant of the dataset), it was decided to use the original unprocessed recordings in BioSemi's .bdf format to ensure better control over data processing and to allow more flexibility in epoching necessary for the predictive task.

2.2.2.1 BIDS data organization

The Python package MNE-BIDS (Appelhoff et al., 2019; Pernet et al., 2019) was used to reorganise .bdf files into BIDS format. BIDS is a standardised way of organising and sharing biosignal data, which allows for better code transfer and replicability in analyses (Gorgolewski et al., 2016). To achieve this, BIDS defines for a given dataset the file formats, file names, directory structure and metadata in a consensual way, thus avoiding arbitrary data organisation. This results in time savings in data curation, a reduction of errors due to misunderstandings, as well as an increased automation of the data analysis workflow (Gorgolewski et al., 2016).

The original .bdf files were first read as *MNE-Python Raw data* structure (based on .fif files) and a series of minimal processing was performed. Firstly, the channel types of the sensors were correctly assigned (e.g. 'EXG1' as 'eog' instead of 'eeg'). Then, for subjects 28 to 32. an empty channel was dropped ('-0' channel) and the

channel marking the stimulus was renamed as 'stim' channel. In addition, 50 Hz was specified as power line frequency (as required by BIDS).

In order to match the event IDs to data, the DEAP database documentation was used (see Table S1). However, challenges were faced in this process. Although the DEAP database has documentation that specifies the correspondence between each event ID and its description, in 9 participants (subject 24 to 32) this correspondence was not documented. According to the DEAP database documentation, the range of IDs events is 1-7 (e.g. start of music video has status code '4'). However, for these 9 subjects event IDs took surprisingly more extreme values (e.g. 1638148). Thus, it was necessary to make a visual inspection of the recording of these participants to be able to infer the link between descriptions and their event IDs. Considering that the sequence of events was given by the experimental design, this encoding could be done accurately. Thus, by plotting the signals and displaying the events as labels in the recordings it was possible to note which ID code corresponds to which event description (e.g. 1638148 ID correspond with the presentation of the stimulus).

Finally, the *write_raw_bids()* MNE BIDS function was used to organise the files according to BIDS-compatible folder structure (see Figure S1).

To check that the data was organised accordingly to the BIDS guidelines, BIDS-Validator was used (Pernet et al., 2019). This platform allows to online validate whether it is used BIDS correctly, and for example check if there are missing data. When counting the events for each subject, it was observed that indeed all subjects were presented with 40 stimuli, with the exception of subject 28 ($N = 37$). As it was not documented that this subject had a reduced number of stimuli presented, this participant was discarded from further subjective arousal decoding analyses (while it was maintained for EDA decoding analyses in which the stimulus information is not of particular relevance).

2.2.2.2 Data segmentation

As continuous recordings were taken for each participant, they contained sections that had to be necessarily discarded from subsequent analyses if we intend to work only with data regarding participants' responses to affective stimuli. These

discarded segments were:

- Segment between the beginning of the signal recording and the first stimulus of the experimental procedure.
- Segments between trials.
- Segment between blocks.
- Segment between the last stimulus of the experimental procedure and the end of the signal recording.

Thus, segments were marked for rejection any time participants' recordings did not correspond to the presentation of the emotional stimuli (nor the fixation cross after the presentation of that stimulus). This selection of segments to reject was done programmatically, by implementing code that uses the original .bdfs files as input and generates as output annotation files (*mne.Annotations* class, with .fif file format) based on the events of each participant.

2.2.2.3 BIDS Pipeline

With the original data organised according to BIDS and with the annotations already appended to the raw files, we proceeded to pre-process the data following the MNE-BIDS-Pipeline (https://mne.tools/mne-bids-pipeline/). What this pipeline provides is a systematic way to analyse BIDS raw data, by directly setting the processing parameters in a configuration file provided by the library. At the times this thesis is being written, this pipeline is still undergoing development. Therefore, it was necessary modify some components of the library to make the pipeline flexible enough to be able to analyse the data from the DEAP Database. In this sense, the pipeline was in the first place intended to analyse M/EEG data. However, as in our case we also needed to process peripheral data (e.g. EDA), the corresponding changes were made to be able to convert and process these signals.

Specifically, the MNE-BIDS-Pipeline was used for data preprocessing (i.e. filtering, artifact rejection and epoching). A minimal preprocessing to physiological

data has been applied. As a first EEG processing step, a 49 Hz low-pass filter was set in this config file. For EDA, a band-pass filter of 0.05 Hz and 5 Hz was implemented as lower and upper limits respectively. Filtering the slower 0.05 Hz frequency waves would remove tonic changes in the EDA signal, while cleaning signals above 5 Hz would avoid artefacts whose source is not given by the nervous system activity (Belfi et al., 2016; Boucsein, 2012; Gerster et al., 2018). Finally, EMG was digitally filtered with a 20 Hz high-pass filter, selected according to published guidelines for this particular psychophysiological technique (De Luca, 1997; Van Boxtel, 2001)

Then, SSP (signal-space projection) vectors were computed to mitigate ocular artifacts in EEG signals (Uusitalo and Ilmoniemi, 1997). To this end, it was necessary to first obtain the peaks of eye blinks in the EOG data to generate epochs around EOG artifact events. After this, the rejection threshold of the EOG artifacts (EEG values for which the presence of EOG artifacts is rejected) was calculated through Autoreject library (Jas et al., 2017). Finally, these SSP vectors were calculated for each subject to deal with these EOG artifacts.

Another variable to specify in the config file were the epoching parameters. In the first instance, continuous 5-second epochs were created, with a 1 s overlap for EEG and EDA data. This epoch length was implemented considering that the latency between the onset of an arousing stimulus and its corresponding EDA response is typically between 1 and 5 s (Boucsein, 2012). Thus, this time window should capture such a response in case it exists. On the other hand, the combined EEG-EMG signal was epoched separately by using a different config file with equivalent parameters. This way, continuous epochs were also obtained, but using a sliding window-approach with 1.5 s windows spaced by 250 ms. These time window was chosen according to previous studies of continuous muscular activity prediction (Sabbagh et al., 2020).

As proposed in the pre-registration document, after epoching the *global* version of Autoreject (Jas et al., 2017) was to compute possible artefacts that were not corrected by SSP vectors. This algorithm automatically selects a threshold for each channel type that, if exceeded at a given epoch, is then excluded from further analysis. However, by applying global version of Autoreject more than a half of the epochs would be discarded for most participants. This entails a major

issue if we consider the amount of data needed for a correct implementation of statistical learning models. In this way, *local* version of Autoreject was applied as an alternative option. This algorithm deals with bad sensors by interpolating this signal with nearby channels. This obviously entails preserving more data, considering it only remove epochs in case of not being able to interpolate a bad signal with surrounding sensors.

2.2.3 Physiological modelling

One of this thesis' aims is to implement EDA prediction models from EEG. However, in order to avoid overfitting, it would entail working directly on this outcome. Thus, we focused first on EMG decoding task (as an unrelated measure). After the predictive model pipeline was optimised with EMG, the same procedure was repeated (with the exact same parameters) for EDA regression.

As such, we first approximated EMG mean and variance from EEG as a weighted sum of spatial information in different frequency bands using a linear regression model. The frequency bands used to perform this approximations were: low (0.1, 1.5 Hz), delta (δ: 1.5 -4Hz), theta (θ: 4-8Hz), alpha (α: 8-15Hz), beta low (β *low*: 15-26Hz), beta high (β *high*: 26-35Hz) and gamma (γ: 35-49Hz). Thus, each epoch of each participant was represented as a between-tsensors covariance matrix of EEG signal. Then, different models were used for the prediction of the outcome. In addition to the *SPoC* and *Riemann* models described above, *diag* and *upper* regression models were also included (see Sabbagh et al., 2020). The *diag* model consists of using only the diagonal of the covariance matrix of sources for the implementation of the predictive model. That is, using the variance (i.e. power) of each sensor as a regression model. On the other hand, the *upper* model consists in using the upper triangular coefficients of covariance matrices, with the off-diagonal terms weighted by a factor$\sqrt{2}$. Furthermore, two versions of the models which include the projection step (i.e. Riemann and SPoC) were created: a full-rank model and a model with the number of components optimised for each participant, considering that previous work has shown an improvement in performance in low-rank models (Sabbagh et al., 2020). Implementations for all these models are

included in the available methods of the *make_filter_bank_regressor()* function of the Coffeine library (https://github.com/dengemann/coffeine).

We, therefore, constrain parameter fitting with an L2 (ridge) penalty that would push the solution towards the principal components that explain more variance. Hyper-parameter search was controlled through grid search (nested cross-validation). To avoid overfitting, predicted physiological signals were generated using 2-fold cross-validation (adapted for time-series to avoid auto-correlation). R^2 was used as the performance metric (i.e. coefficient of determination), considering there was no particular interest in interpreting the results in terms of the units of the physiological variable.

2.3 Predictive modelling of self-reported arousal

The predicted EDA, alongside the observed EDA, were subjected to a linear mixed-effects model with stimulus and subject as random effects. Uncertainty estimates were obtained from non-parametric bootstrapping of the entire process with 2000 iterations Our proposed model had ”observed EDA” and ”EEG-enriched EDA” as independent variables for the subjective arousal regression analysis. Thus, finding an additive contribution of the predicted EDA to the observed EDA would mean that the predicted EDA would have information that is not redundant for the arousal regression task in comparison to the observed EDA.

3 Results

3.1 Predictive modelling from EEG

3.1.1 Epochs for predictive modelling

After completing the specified preprocessing steps, the median number of epochs obtained for further predictive EEG-EMG analysis was 8979.5 (*IQR* = 6950 - 9913.75) epochs, while the median number of epochs for EEG-EDA analysis was 499.5 (*IQR* = 377 - 592) epochs. The distribution of the final number of epochs

per subject for both EEG-EMG and EEG-EDA predictive modelling tasks can be observed in Figure S2.

3.1.2 EMG decoding

Preliminary predictions of the EMG mean and variance were obtained for each subject. Then, performance metrics at the level of each participant for each model evaluated were considered. The observed results showed that there were differences in the performance of the EMG predictions dependent on the descriptive statistic used as a target for the models' implementation (see Figure S3). On the one hand, no model could outperform the prediction by chance in the EMG *mean* prediction task. On the other hand, for the EMG *variance* prediction task, differences in performance were found depending on the model used. Thus, Riemann models (both its full-rank and the component optimised version) and *upper* model achieved performances that did not exceed the random prediction. However, by implementing *diag* and SPoC models (both in their full-rank and component optimised versions) it was found a performance that exceeded the chance performance threshold in most of the analysed subjects. While at the particular individual predictions, performances close to $R^2 = 0.5$ (e.g. Figure S4) have been found, it is important to note that the median prediction across subjects did not exceed $R^2 = 0$ for any model. However, it is worth noting that the main purpose of these predictions was to optimise the predictive pipeline to avoid overfitting during the EDA decoding task.

3.1.3 EDA decoding

EDA mean and variance predictions were also obtained for each subject, in the same way to that reported for the EMG decoding task (see Figure 6). In the case of predictive modelling of EDA, similar results were found for the prediction of the mean of this signal compared to what was found in EMG: no model performed better than prediction by chance.

However, for the prediction of EDA variance, the prediction with the SPoC and diag models not only outperformed the prediction by chance for most of

the participants, but also the median of the coefficient of determination of these predictions of these subjects was above zero: $R^2 = 0.11$, $R^2 = 0.23$ and $R^2 = 0.355$ for SPoC full ranked, SPoC component optimised and diag models, respectively.

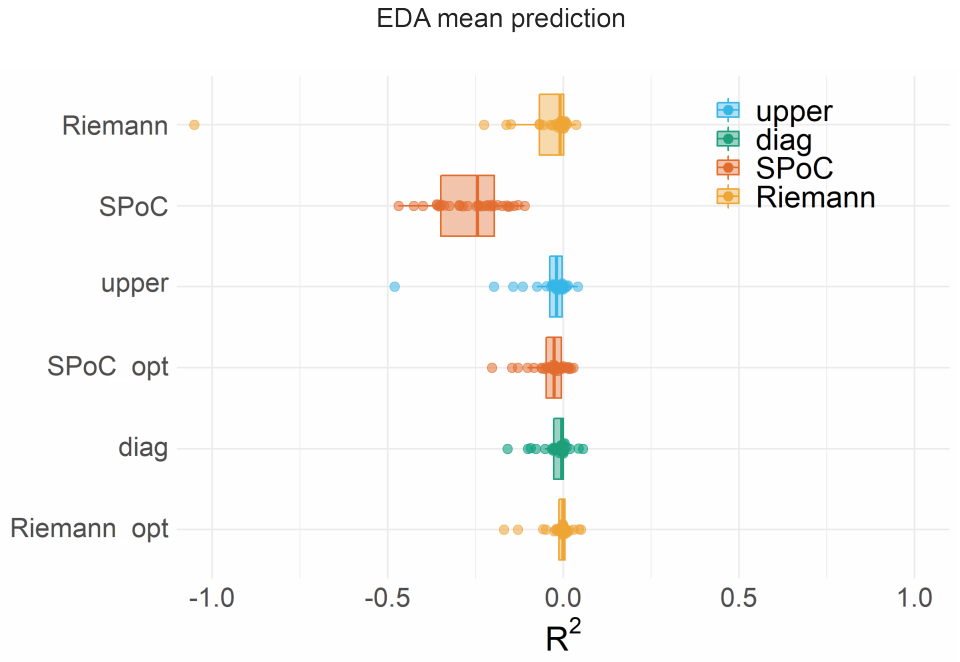

Figure 6. Model comparison in EDA mean prediction across subjects. The different models to be compared are represented on the Y-axis of each subfigure, while performance is represented on the X-axis. Each type of model has a different colour in the figure. Thus, dots represent the average expected out of sample performance (R^2) of the 2-fold cross validation of each participant. The distribution of participants' performance on each model is summarized by boxplots.

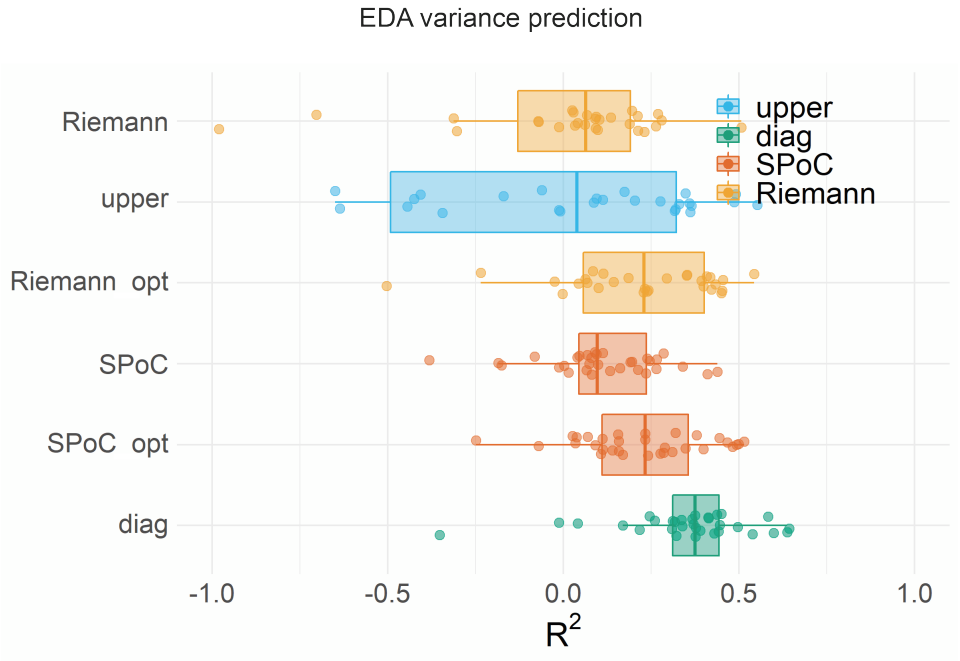

Figure 6. Model comparison in EDA variance prediction across subjects. The different models to be compared are represented on the Y-axis of each subfigure, while performance is represented on the X-axis. Each type of model has a different colour in the figure. Thus, dots represent the average expected out of sample performance (R^2) of the 2-fold cross validation of each participant. The distribution of participants' performance on each model is summarized by boxplots.

In Figure 7, the prediction of single-subject variance EDA for subjects 15 and 16 can be observed. Thus, it is possible to note that the individual predictions for *diag* model were $R^2 = 0.599$ and $R^2 = 0.643$ respectively. To inspect the performance of all participants in EDA predictive modelling task with *diag* models, see Figure S5.

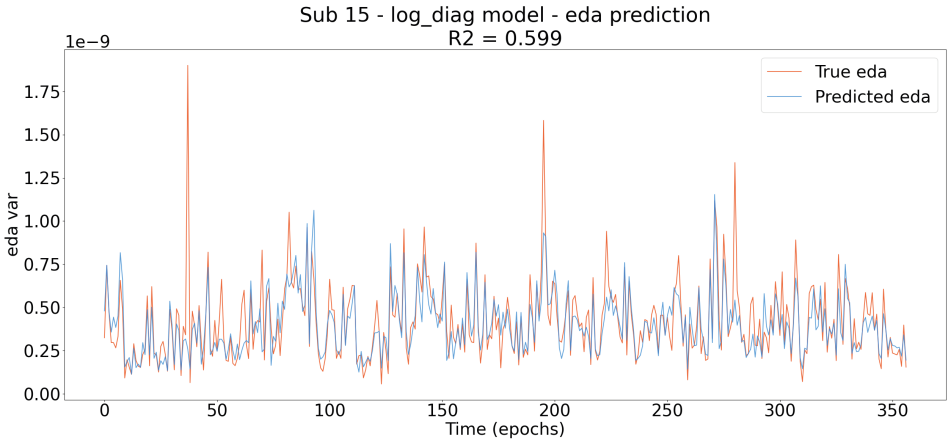

Figure 7. Continuous EDA variance decoding with *diag* model for subject 15. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

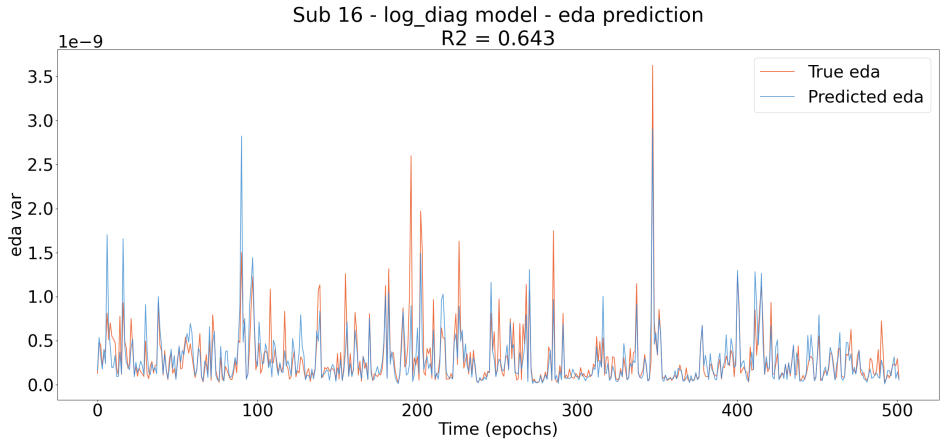

Figure 7. Continuous EDA variance decoding with *diag* model for subject 16. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

3.1.3.1 Components of EDA decoding models

Once the corresponding predictive analyses with SPoC were applied to each participant, it was possible to identify which components were most correlated with the modeled outcome. In this way, visualisations of the contributions of the six principal components detected in the EDA variance decoding task were made (see Figure 8 and Figure S6).

At a general level, global patterns of activity were observed in most of the components detected by SPoC. In addition, it is possible to observe the contribution of EOG artefacts (e.g. first two SPoC components of subject 4, Figure S6).

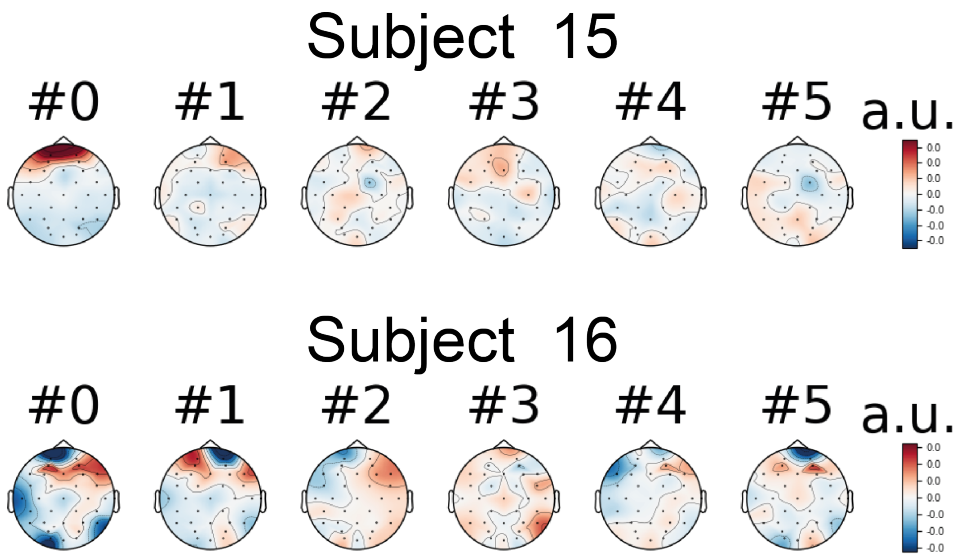

Figure 8. Plot of the contribution of the five first detected components in EDA variance decoding task for subject 15 and 16. Colours indicate the contribution of that sensor space to the SPoC patterns represented. Distributed patterns of activation can be observed in the detected components, in addition to components related to ocular artefacts of the participants.

	Linear model (no random effects)
(Intercept)	−0.01
	(0.03)
True EDA	−0.23**
	(0.09)
Predicted EDA	0.09
	(0.09)
R^2	0.02
Adj. R^2	0.02
Num. obs.	1170

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 1. Subjective arousal lineal model

3.2 Predictive modelling of self-reported subjective arousal

In the first instance, a simple linear model was computed by considering the prediction of participants' self-reported subjective arousal as a function of True EDA and predicted EDA:

$$Arousal = \beta_0 + \beta_1 True\,EDA + \beta_2 Predicted\,EDA + \varepsilon \tag{1}$$

The variables were Z-scored standardised prior to the analysis, to ensure that the estimated coefficients had the same scale. Thus, the following results were obtained (see Table 1).

In the first instance, True EDA would be a significant predictor of participants' self-reported arousal (β= -0.232, $p < 0.01$, CI = [-0.04, -0.06]), while Predicted EDA was not a significant predictor of arousal (β= 0.088 , $p > 0.05$, CI= [-0.08, 0.26]).The overall model fit was R^2 = 0.02.

However, this model would assume independence in our data. On the contrary, data from 31 different subjects were recorded, who were exposed to 40 emotional stimuli. Thus, it is reasonable to consider that there is a correlation between the observations dependent on the subject and the stimulus content. Consequently, a linear mixed-effects model with stimulus and subject as random effects was

performed ($Arousal_{ij}$ denote the i-th subjective arousal report to a particular stimulus of the j-th participant) :

$$Arousal_{ij} = \beta_0 + \beta_1 True\,EDA_{ij} + \beta_2 Predicted\,EDA_{ij} + u_i + v_j + \varepsilon_{ij} \qquad (2)$$

In this case, Predicted and True EDA were set as fixed effect variables (i.e. variables that are expected to have an effect on subjective arousal). Then, we included random intercepts in this model, allowing the mean response to depend on subject and stimulus. This is so considering that all participants watched the same music clips.

Thus, the following results were obtained from this model (see Table 2)

Again, it is possible to observe that True EDA is a significant predictor of participants' self-reported arousal (β= -0.175, $p < 0.05$, CI = [-0.33, -0.01]), while the inclusion of EDA prediction would not seem to have additive effects in explaining this outcome (β= 0.121, $p > 0.05$, CI= [-0.04, 0.27]).

However, it is important to note that moderate collinearity has been found between the True EDA and Predicted EDA variables (Variance Inflation Factor [VIF] = 9.03), which could be considered a non-tolerable correlation of predictors in linear models (Ringle et al., 2015).

	Nested linear mixed-effects model
(Intercept)	−0.01
	(0.09)
True EDA	−0.17*
	(0.08)
Predicted EDA	0.12
	(0.08)
AIC	3161.72
BIC	3192.11
Log Likelihood	−1574.86
Num. obs.	1170
Num. groups: stim	40
Num. groups: subject	31
Var: Stimulus(intercept)	0.14
Var: Subject (Intercept)	0.10
Var: Residual	0.77

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 2. Subjective arousal nested linear mixed-effects model

4 Discussion

4.1 Summary of main results

The present thesis aimed to perform predictive modelling of EDA from EEG using a public database. At the same time, it was proposed that such predictions could be useful for predicting subjective arousal in response to emotional experiences.

To achieve the former goal, first it was required to define summary measures of the EDA signal to be predicted. As this was an exploratory project, considering that there are no other studies in the literature that have addressed this task, it was decided to use the mean and the variance of the signal as the outcome measure.

After fine-tuning the pipeline with the EMG signal, the first goal was achieved. From the present work it was possible to implement a regression pipeline to predict the EDA at the participants' event-level with the data obtained from EEG recordings. These predictions achieved even good performances at the level of each participant (above $R^2 = 0.3$ in 24 of the 32 subjects), even controlling for various sources of error. One of the particularities of the results found is that the *diag* model was the one with the best performance for the prediction of this signal, outperforming the Riemann and SPoC methods. By observing the principal components of the latter method it is possible to see global patterns of activation across the sensors.

However, a linear mixed-effect model including both *Predicted EDA* and *True EDA* for the subjective arousal regression did not find the expected additive effects. That is, despite having achieved good performance in the prediction of EDA, this predicted signal would not contribute to the prediction of the participants' self-reported arousal as hypothesised in the present thesis.

4.2 From neurophysiology and biophysics to statistical modelling

EMG signal was particularly chosen as a first measure to develop a prediction pipeline considering that it has been previously used in MEG predictive tasks with regression models analogous to those used in this thesis (see Sabbagh et al. (2020).

However, in this previous study data provided by Schoffelen et al., 2011, whose task specifically involved participants repeatedly contracting and extending their hands, was used. Thus, we can see that their procedure was designed to maximise the signal-to-noise ratio (SNR) in the presented task. This contrasts with the predictive task performed in the present work, where the recording of zygomatic muscle activity is one of several measures obtained while participants observe music videos. On the other hand, the length of epochs used may be useful for the prediction of motor tasks, while it should be optimised to tackle EMG prediction during affective tasks. Furthermore, while there is evidence in the literature of the relationship between such EMG activity and the valence component of affective states, its relationship to arousal is far from conclusive (Sato et al., 2020). Therefore, the achievement of positive coefficients of determination at least in some participants may be the starting point for further research focusing specifically on the prediction of this signal from EEG in relation to arousal.

Regarding the prediction of EDA signals, one of the remarkable results is the performance improvement when predicting the variance of the signal instead of the mean. This result can be interpreted based on the properties of the peripheral pathways related to the generation of electrodermal activity (see Greco, Valenza, and Scilingo, 2016). The eccrine glands in charge of sweat production are innervated by sudomotor nerves. The modulation of EDA is thus preceded by bursts of these nerves that control these glands. These bursts are temporary discrete episodes (Greco, Valenza, Lanata, et al., 2015), which are often associated with the phasic component of the EDA signal, such as skin conductance responses (SCRs). SCRs are at least in essence a measure of variability, considering that traditional methods to obtain an SCR is to measure the range between the minimum and maximum point of a window of interest. Thus, it is clear that this measure is more related to the variance than to the mean of the signal, the latter being more associated with a tonic component of the EDA (Boucsein, 2012). Consequently, it is to be expected that if we are measuring the response to particular events (i.e. parts of the music clips), this would be more associated with phasic than tonic issues, and therefore the variance may be better predicted than the mean EDA signal.

On the other hand, among the predictive models of EDA variance, the one that

achieved the best performance was the diag model. In order to interpret this result, it is necessary to consider biophysical mechanisms related to the EEG signal (Hari and Puce, 2017). It is assumed that there are multiple cortical generators, each of which projects different electromagnetics fields on the EEG sensors (Sabbagh et al., 2020). These generators would be combined by an unknown function which is additive. This is justified, for example, by the existence of independent components on the brain (Hyvärinen and Oja, 2000) or by the fact that this neurons' activity is a summing process (e.g. the amplitude of the EEG is proportional to the number of neurons participating in the activity). Thus, lineal models in combination with spatial filtering or Riemann modelling is a way to summarize the EEG signal with concerning the spatial dimension, to recombine the sensors to enhance SNR of the dynamics. Considering this, it is striking that in the first instance the diag model, which does not consider the covariance between sensors, performed better than SPoC and Riemann methods.

A better performance in the diag model would indicate that the extracted information is more univariate, and the intercorrelation between channels is less important. This result may be consistent with a diffuse generator of EEG signal, and not necessarily cortical. This pattern has been observed for example in the case of the prediction of states of consciousness in patients with consciousness disorders (Engemann, Raimondo, et al., 2018). In such patients a change of global power of the signal, particularly in the alpha and theta band, is a better predictor of the state of consciousness than the connectivity between electrodes. Is it possible that this is also the case for the arousal phenomenon? A partial answer to this question can be obtained by observing the representations of the components found through the SPoC method (see Figure S6). From these visualisations it is possible to note that there are indeed no spatially defined components, but rather sparse activity along the cortex (in addition to some ocular artefacts). This is consistent with the idea that the overall information on the arousal is not encoded in specific cortical sources. This is in contrast to the components related to EMG prediction in relation to a hand movement prediction task, as was the case for Sabbagh et al. (2020), which has shown higher performance in the SPoC and Riemann models than in diag models. Another possible explanation, closer to the knowledge concerning the sources of

EDA (see 1.1.4), is that the source of the EEG signal in relation to arousal is mainly subcortical. This being the case, it is consistent that the diag model performs better than the other candidate models, and it is also expected to observe such patterns in the SPoC principal component representations.

However, these results contrast at least partially with those obtained by Hofmann et al., 2020, who found decreased activity in alpha power predominantly in parieto-occipital electrodes. However, it is important to note that in this case the outcome was subjective arousal and not EDA. Nevertheless, to assess whether there is indeed a contribution of alpha activity in the predictive models of EDA, it would be possible to repeat the analyses after running a Notch filter (i.e. band-stop filter) to alpha frequency band. Thus, if we find a reduction in prediction performance compared to using such a filter in other frequency bands, this would indeed indicate the possible contribution of alpha frequencies in EDA predictive modelling. Certainly one aspect to highlight is that through the use of linear models, such as those used in this thesis, it is possible to go back to the feature space and make representations like those obtained with SPoC. This allows us to interpret the phenomenon and draw appropriate conclusions, considering both the characteristics of the model implemented (i.e. assumptions of the method used) and the parameters related to the features used to make those predictions. This is one of the main advantages of linear models in comparison with the non-linear models (e.g. long short term memory neural networks), with which it is possible to increase the performance of the model but often at the expence of losing in the interpretability (Vellido, 2020).

4.3 Limitations of predicted EDA proxy measure approach

The negative result obtained regarding the EDA proxy measures can be commented from different angles. First, in this type of models (as generally in frequentist statistics) the sample size is important to increase the probability of finding results that reject the null hypothesis in case it is false (power size). In this case, when implementing a mixed model with 40 presented stimuli, if we set $\alpha = 0.05$ and assume a moderate effect size (0.3) we would need a sample size of 376 participants to have a 1- $\beta = 0.8$ (Snijders, 2005). In our case we had only 31 subjects on record

(considering that one subject was discarded as part of the pre-processing). Thus, it is very likely that even if there was an effect we could not see it because of the low statistical power. Therefore, new databases with larger samples sized in affective arousal tasks may help to understand whether this approach could indeed be valid to improve the predictions of these signals.

On the other hand, a strong collinearity has been observed between the True and Predicted EDA variables. This may have made the obtained results difficult to interpret. One way to address this issue could be through a add a further level of complexity to the statistical model. As strong collinearity has been found when averaging the variance over many time points, one possibility to deal with this is could be to perform a linear mixed-effects model using a liner mixed model where all time points are concatenated across all subjects. In this way, instead of having one data point per stimulus per subject, we would have n data points (where n refers to the number of EEG-EDA epochs corresponding to the presentation of a given stimulus). Thus, it is possible that the collinearity is reduced using this approach, and that the models could be correctly implemented . Moreover, this way it is possible to evaluate the possibility analyzing subject by subject, considering that perhaps additive effects occur in some subjects but not others.

Another option would be to increase the number of trials for each participant. It should be also considered the possibility of implementing individual models for each subject that allow general predictions at the event-level, as was the approach used in the first step of the present thesis (i.e. prediction of EDA from EEG).

If the classical form of self-reporting used in affective computing tasks is used (i.e. observing an emotional stimulus and then giving a single self-report after its presentation as a summary of the entire affective state during that stimulus) some methodological concerns would be implied. To begin with, if it is intended to use videos as the ones used in out procedure (1-minute videoclips), this would require the subject to be evaluated for several hours in order to have a total of about 100 stimuli presented. Another possible option would be to show shorter stimuli, such as affective images. However, this type of stimuli lose strength for arousal generation and therefore would reduce the SNR that we seek to maximise. At the same time, exposing the subject to a very long affective task increases the chances of habituation

during the task. Alternatively, we propose the possibility of obtaining a continuous report of the participant's arousal, in a similar way as physiological signals are obtained continuously. This would imply that the participant's affective state can be continuously monitored, as was previously done for example in Hofmann et al. (2020). In this study, continuous arousal recordings were obtained from participants, which used a dial to manipulate a scale dimension with values from 0 to 100 (lower to higher arousal values, respectively). This type of continuous measurement was also registered in DECAF public database (Abadi et al., 2015), which contains MEG, EMG and ECG data (among other measures). This dataset can potentially be used with the same approach used in the present work to understand coupling dynamics between central nervous system signals, peripheral nervous system signals, and behavioural variables.

Another issue to consider is whether the use of the EDA variance (not only the predicted signal but also the true signal) is a good predictor of arousal. At first, this variance was used as an exploratory outcome summary during EDA prediction from EEG covariance matrices. However, in the future it may be necessary to extend this approach by using more physiologically relevant electrodermal activity signals. For example, extracting the phasic signal from the EDA at each epoch with signal deconvolution algorithms (Benedek and Kaernbach, 2010; Greco, Valenza, Lanata, et al., 2015), and summarising that measure from some of its characteristics. For example, obtain the average amplitude of SCRs, pick count during the stimulus presentation, or the area under the curve of SCRs as a summary of the signal (see Shukla et al., 2019). Perhaps by using more physiologically related aspects of the arousal signal such as SCRs it is possible to achieve the additive effects hypothesised in the present work. It is clear that a better understanding of the coupling mechanisms between central and peripheral nervous system signals is a task that could lead to the effective use of the predicted signals as a proxy measure of arousal.

On the other hand, it could be that the affective task proposed in the database used during the present thesis is also not maximising the SNR for EDA prediction, nor subjective arousal decoding. Although in the DEAP database stimuli with more extreme values in relation to both the valence and arousal dimensions, were selected

by the authors, it is possible that the very nature of music videos as stimuli does not guarantee a good separation between low and high arousal states. For example, the protocol recently used by Hofmann et al. (2020) involves the use of immersive stimulation from VR during the simulation of roller coaster experiences. As it can be seen, this type of task was specifically designed to maximise the difference in the arousal dimension in the participants' self-reports, which could result in more appropriate conditions to then apply proper statistical learning algorithms for the prediction of the subjective reported arousal. This type of immersive environment, which has also been used in clinical contexts for exposure to anxiety disorders (Carl et al., 2019), may have the necessary qualities to become a gold standard for the elicitation of affective states in future databases, considering the flexibility of these environments, the experimental control allowed, and the engaging context for participants that this technology allows (Boeldt et al., 2019).

4.4 Trade-off between guarantee reproducibility and ensuring a better SNR

One of the aims of this thesis was to carry out the corresponding predictive tasks attempting to maximise transparency in the procedures for obtaining, organizing, preprocessing and analysing data. This type of approach was carried out so that they could be effectively replicated by other researchers interested in the subject.

We worked with a database that can be accessed for research purposes (upon request to the authors). The data were also systematised according to the BIDS organisation using MNE BIDS Python package. The use of this standard allows the replicability of the results, avoiding arbitrary decisions regarding the type of artifacts used, the name of the artifacts, metadata information and other issues related to data management. At the same time, a programmatic approach to data preprocessing was used, employing MNE-BIDS-Pipeline library. Although in our case it was necessary to adapt the pipeline to make it sufficiently flexible for the tasks we wanted to address, considering that the package is still under development, the use of such pipeline through a configuration file allows us to systematise the steps in the EEG processing. This would avoid arbitrary decisions in the pre-processing of the data

that could lead to different results.

Using this approach also has some drawbacks. For example, it was thought at first to include manual annotations of the participants' respiration artifacts to the signal, with the idea that these could then be rejected when epoching. Although they were computed in the present work (and can be accessed in the repository available 5), they were not used to compute the epochs of the predictive motions in order to guarantee a programmatic approach in the data processing. However, this decision may be to the detriment of better signal quality.

Another decision that may have influenced the quality of the data is the use of SSP vectors for the treatment of eye artefacts. These vectors were computed related to eye activity were computed and then automatically removed from the EEG signal. However, although this method is robust for the detection of such artefacts (Sabbagh et al., 2020), when representing the first components of the SPoC method it is possible to observe the presence of some EOG components in the data. This would mean that this method allows some ocular artefacts to pass through to the EEG signals (e.g. first two SPoC components of subject 04, Figure S6), which could in part be a limitation of the present work. An alternative to this would be to compute the independent sources of the signal with ICA to manually treat the artefacts. Nevertheless, this would consist of a manual approach, possibly improving the quality of the signal obtained. In this way, it would be possible to assess whether the results found are sustained considering the removal of clearly ocular components from the EEG data. However, it is also possible that eye artifacts may have information to add the link between behavior and arousal. Hence, if it is an artifact depends on proposed goals.

4.5 Perspectives

4.5.1 Generalisability of results

The database used contains recordings of the participants in two separate locations (subjects 1 to 22 were evaluated in Twente, while 23 to 32 in Geneva). Thus, predicting the electrodermal activity for the subjects in both locations leads

to a certain degree of generalisability of the results. However, in order to be more confident that the results found are generalisable and not dependent on the database used, the same analyses described in the present thesis could be applied in the future to other datasets. This task can be facilitated thanks to the support provided by the MNE BIDS Pipeline library, especially in relation to the specification of common steps in data pre-processing. Thus, the MAHNOB-HCI (Soleymani et al., 2011) database, which contains the emotional responses of users to emotional movies, images and short videos, could be used to evaluate the EEG-EDA relationship. On the other hand, it could be evaluated whether it would indeed be possible to achieve consistent performances in the modelling of the EEG-EMG relationship by evaluating the EEG-EMG relationship on another dataset. However, this is a challenge, considering the non-existence of another dataset that brings together both measures in an emotional task. Therefore, we will use the DECAF dataset (Abadi et al., 2015) that gathers the magnetoencephalography (MEG) as well as the EMG signal of participants while watching emotional videos. Considering the intimate relationship between EEG and MEG signals (see Hari and Puce, 2017), it would be possible to make comparisons (and assess possible generalisability) with EEG and EMG analyses performed from the DEAP database.

At the same time, it would be of interest to address other psychophysiological signals with this predictive modelling of arousal approach. Is it possible to achieve a multidimensional mapping of arousal by relating different peripheral signals with cortical components? To tackle this question, the EEG-HRV relationship could be modeled, which has been extensively studied for its contribution to the explanation of the arousal phenomenon (Nardelli, Greco, et al., 2018; Nardelli, Valenza, et al., 2015). There are public databases that allow us to address this issue, considering that this measure can be extracted from the electrocardiogram (ECG), which is among the measures evaluated in the MAHNOB-HCI database.

4.5.2 Clinical implications of predictive modelling of arousal

A better understanding of the coupling mechanisms between peripheral and central nervous system measures of arousal may not only have relevance at a basic

level, but may also have clinical implications.

Although arousal is a dimension that has historically had strong influences at the clinical level (Hoehn-Saric and McLeod, 2000; Van den Burg et al., 1992), and since the establishment of arousal as one of the main dimensions across mental disorders in the RDoCs (First, 2012) it has gained even more relevance in the state of the art. This type of approach has the quality of being transdiagnostic, focusing on point processes that may be common to different conditions across different levels of organization, from the molecular level to the self-report (Carcone and Ruocco, 2017).

Among them, the relationship of arousal in relation to anxiety disorders is evident. In addition, the relationship of this component in depression has been investigated in relation to the lack of motivation as a factor of maintenance of this disorder (Cléry-Melin et al., 2011). Previous studies have specifically investigated the relationship of EEG with reports of depressive symptomatology, finding associations of EEG vigilance features (e.g. slower arousal decline) with different range values on a major depressive scale (Schmidt et al., 2017). Is it possible from predictive modelling of EDA to think of possible markers of this disorder that bring together EEG information with other peripheral measures?

Finally, while the efficacy of pharmacological treatment is known (at least at the level of symptomatic reduction), there is a gap in the literature concerning the improvement of the understanding of the mechanisms that enable such therapeutic action. This becomes particularly relevant in relation to why selective serotonin reuptake inhibitors (SSRI) also improve the motivational and arousal components in patients (Cowen and Browning, 2015). Olbrich et al. (2016) investigated how certain central and peripheral nervous system patterns (EEG and HRV features, respectively) could predict the response to antidepressant medication. Is it possible to gain a better understanding of the pharmacodynamics of these molecules by modelling the relationship between peripheral and central measures with EDA? Could features that bring together characteristics of both measures, such as the proxy measures used in the present thesis, be useful in such a predictive task? Again, these questions could have implications in relation to any clinical diagnosis in which arousal is a domain of relevance.

4.6 Conclusion

The present thesis constitutes the first attempt to use a two-step approach for predictive modelling of arousal. This methodology consists of predicting a physiological measure and then using that predicted signal as a proxy for subjective arousal.

In the first instance, it is important to highlight that positive results were found for the prediction of EDA. This means it is indeed possible to predict the EDA from EEG signals. However, the signals chosen as a summary of EDA across the participants' epochs do not seem to have a solid contribution to the explanation of EDA. This, in addition to other limitations of the database used, may have resulted in the absence of an additive contribution of the predicted signal to the EEG signal in explaining the participants' subjective arousal.

Thus, considering the programmatic approach of the processing steps used in the methodology, it remains open the possibility of exploring other measures to summarize the EDA signal, as well as other psychophysiological methods, to achieve a better understanding of the arousal phenomenon.

5 Code Availability

The source code of the implementations used to compute organization, processing and statistical analyses can be obtained from https://github.com/tomdamelio/arousal_decoding

6 References

Abadi, M. K., Subramanian, R., Kia, S. M., Avesani, P., Patras, I., & Sebe, N. (2015). Decaf: Meg-based multimodal database for decoding affective physiological responses. *IEEE Transactions on Affective Computing*, *6*(3), 209–222.

Appelhoff, S., Sanderson, M., Brooks, T. L., van Vliet, M., Quentin, R., Holdgraf, C., Chaumon, M., Mikulan, E., Tavabi, K., Höchenberger, R. et al. (2019). Mne-bids: Organizing electrophysiological data into the bids format and facilitating their analysis. *The Journal of Open Source Software*, *4*(44).

Arevalillo-Herráez, M., Cobos, M., Roger, S., & Garcıa-Pineda, M. (2019). Combining inter-subject modeling with a subject-based data transformation to improve affect recognition from eeg signals. *Sensors*, *19*(13), 2999.

Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on psychological science*, *1*(1), 28–58.

Barrett, L. F., & Fossum, T. (2001). Mental representations of affect knowledge. *Cognition & emotion*, *15*(3), 333–363.

Bates, D., Sarkar, D., Bates, M. D., & Matrix, L. (2007). The lme4 package. *R package version*, *2*(1), 74.

Belfi, A. M., Chen, K.-H., Schneider, B., & Tranel, D. (2016). Neurological damage disrupts normal sex differences in psychophysiological responsiveness to music. *Psychophysiology*, *53*(1), 14–20.

Benedek, M., & Kaernbach, C. (2010). Decomposition of skin conductance data by means of nonnegative deconvolution. *psychophysiology*, *47*(4), 647–658.

Boeldt, D., McMahon, E., McFaul, M., & Greenleaf, W. (2019). Using virtual reality exposure therapy to enhance treatment of anxiety disorders: Identifying areas of clinical adoption and potential obstacles. *Frontiers in psychiatry*, *10*, 773.

Boucsein, W. (2012). *Electrodermal activity*. Springer Science & Business Media.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, *25*(1), 49–59.

Campos-Rodrıguez, R., Godınez-Victoria, M., Abarca-Rojano, E., Pacheco-Yepez, J., Reyna-Garfias, H., Barbosa-Cabrera, R. E., & Drago-Serrano, M. E. (2013). Stress modulates intestinal secretory immunoglobulin a. *Frontiers in integrative neuroscience*, *7*, 86.

Carcone, D., & Ruocco, A. C. (2017). Six years of research on the national institute of mental health's research domain criteria (rdoc) initiative: A systematic review. *Frontiers in cellular neuroscience*, *11*, 46.

Carl, E., Stein, A. T., Levihn-Coon, A., Pogue, J. R., Rothbaum, B., Emmelkamp, P., Asmundson, G. J., Carlbring, P., & Powers, M. B. (2019). Virtual reality exposure therapy for anxiety and related disorders: A meta-analysis of randomized controlled trials. *Journal of anxiety disorders*, *61*, 27–36.

Cisler, J. M., & Koster, E. H. (2010). Mechanisms of attentional biases towards threat in anxiety disorders: An integrative review. *Clinical psychology review*, *30*(2), 203–216.

Cléry-Melin, M.-L., Schmidt, L., Lafargue, G., Baup, N., Fossati, P., & Pessiglione, M. (2011). Why don't you try harder? an investigation of effort production in major depression. *PloS one*, *6*(8), e23178.

Colibazzi, T., Posner, J., Wang, Z., Gorman, D., Gerber, A., Yu, S., Zhu, H., Kangarlu, A., Duan, Y., Russell, J. A. et al. (2010). Neural systems subserving valence and arousal during the experience of induced emotions. *Emotion*, *10*(3), 377.

Cool, J., & Zappetti, D. (2019). The physiology of stress. *Medical student well-being* (pp. 1–15). Springer.

Cowen, P. J., & Browning, M. (2015). What has serotonin to do with depression? https://onlinelibrary.wiley.com/doi/10.1002/wps.20229

Dadi, K., Varoquaux, G., Houenou, J., Bzdok, D., Thirion, B., & Engemann, D. (2020). Beyond brain age: Empirically-derived proxy measures of mental health.

Dähne, S., Meinecke, F. C., Haufe, S., Höhne, J., Tangermann, M., Müller, K.-R., & Nikulin, V. V. (2014). Spoc: A novel framework for relating the amplitude of neuronal oscillations to behaviorally relevant parameters. *NeuroImage*, *86*, 111–122.

De Luca, C. J. (1997). The use of surface electromyography in biomechanics. *Journal of applied biomechanics*, *13*(2), 135–163.

Ekman, P. (1992). Are there basic emotions?

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology*, *17*(2), 124.

Engemann, D. A., Kozynets, O., Sabbagh, D., Lemaıtre, G., Varoquaux, G., Liem, F., & Gramfort, A. (2020). Combining magnetoencephalography with magnetic resonance imaging enhances learning of surrogate-biomarkers. *Elife*, *9*, e54055.

Engemann, D. A., Raimondo, F., King, J.-R., Rohaut, B., Louppe, G., Faugeras, F., Annen, J., Cassol, H., Gosseries, O., Fernandez-Slezak, D. et al. (2018). Robust eeg-based cross-site and cross-protocol classification of states of consciousness. *Brain*, *141*(11), 3179–3192.

First, M. B. (2012). The national institute of mental health research domain criteria (rdoc) project: Moving towards a neuroscience-based diagnostic classification in psychiatry. *Philosophical issues in psychiatry II: Nosology*, *12*, 18.

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences*, *115*(27), E6106–E6115.

Fuller, P., Sherman, D., Pedersen, N. P., Saper, C. B., & Lu, J. (2011). Reassessment of the structural basis of the ascending arousal system. *Journal of Comparative Neurology*, *519*(5), 933–956.

Gerster, S., Namer, B., Elam, M., & Bach, D. R. (2018). Testing a linear time invariant model for skin conductance responses by intraneural recording and stimulation. *Psychophysiology*, *55*(2), e12986.

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O. et al. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific data*, *3*(1), 1–9.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L. et al. (2013). Meg and eeg data analysis with mne-python. *Frontiers in neuroscience*, *7*, 267.

Greco, A., Valenza, G., Citi, L., & Scilingo, E. P. (2016). Arousal and valence recognition of affective sounds based on electrodermal activity. *IEEE Sensors Journal*, *17*(3), 716–725.

Greco, A., Valenza, G., Lanata, A., Scilingo, E. P., & Citi, L. (2015). Cvxeda: A convex optimization approach to electrodermal activity processing. *IEEE Transactions on Biomedical Engineering*, *63*(4), 797–804.

Greco, A., Valenza, G., & Scilingo, E. P. (2016). *Advances in electrodermal activity processing with applications for mental health*. Springer.

Hadley, W. (2016). *Ggplot2: Elegrant graphics for data analysis*. Springer.

Halász, P., Terzano, M., Parrino, L., & Bódizs, R. (2004). The nature of arousal in sleep. *Journal of sleep research*, *13*(1), 1–23.

Hari, R., & Puce, A. (2017). *Meg-eeg primer*. Oxford University Press.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hess, U., & Thibault, P. (2009). Darwin and emotion expression. *American Psychologist*, *64*(2), 120.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Hoehn-Saric, R., & McLeod, D. R. (2000). Anxiety and arousal: Physiological changes and their perception. *Journal of affective disorders*, *61*(3), 217–224.

Hofmann, S. M., Klotzsche, F., Mariola, A., Nikulin, V. V., Villringer, A., & Gaebler, M. (2020). Decoding subjective emotional arousal from eeg during an immersive virtual reality experience. *bioRxiv*.

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural networks*, *13*(4-5), 411–430.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F., & Gramfort, A. (2017). Autoreject: Automated artifact rejection for meg and eeg data. *NeuroImage*, *159*, 417–429.

Kim, J., & André, E. (2008). Emotion recognition based on physiological changes in music listening. *IEEE transactions on pattern analysis and machine intelligence*, *30*(12), 2067–2083.

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., & Patras, I. (2011). Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, *3*(1), 18–31.

Kragel, P. A., & LaBar, K. S. (2016). Decoding the nature of emotion in the brain. *Trends in cognitive sciences*, *20*(6), 444–455.

Kukolja, J., Schläpfer, T. E., Keysers, C., Klingmüller, D., Maier, W., Fink, G. R., & Hurlemann, R. (2008). Modeling a negative response bias in the human amygdala by noradrenergic–glucocorticoid interactions. *Journal of Neuroscience*, *28*(48), 12868–12876.

LeDoux, J. (2012). Rethinking the emotional brain. *Neuron*, *73*(4), 653–676.

LeDoux, J. E., Iwata, J., Cicchetti, P., & Reis, D. (1988). Different projections of the central amygdaloid nucleus mediate autonomic and behavioral correlates of conditioned fear. *Journal of Neuroscience*, *8*(7), 2517–2529.

LeGates, T. A., Fernandez, D. C., & Hattar, S. (2014). Light as a central modulator of circadian rhythms, sleep and affect. *Nature Reviews Neuroscience*, *15*(7), 443–454.

Mangina, C. A., & Beuzeron-Mangina, J. H. (1996). Direct electrical stimulation of specific human brain structures and bilateral electrodermal activity. *International Journal of Psychophysiology*, *22*(1-2), 1–8.

Meinel, A., Castaño-Candamil, S., Reis, J., & Tangermann, M. (2016). Pre-trial eeg-based single-trial motor performance prediction to enhance neuroergonomics for a hand force task. *Frontiers in Human Neuroscience*, *10*, 170.

Merali, Z., McIntosh, J., Kent, P., Michaud, D., & Anisman, H. (1998). Aversive and appetitive events evoke the release of corticotropin-releasing hormone and bombesin-like peptides at the central nucleus of the amygdala. *Journal of Neuroscience*, *18*(12), 4758–4766.

Nardelli, M., Greco, A., Valenza, G., Lanata, A., Bailón, R., & Scilingo, E. P. (2018). A multiclass arousal recognition using hrv nonlinear analysis and affective images. *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 392–395.

Nardelli, M., Valenza, G., Greco, A., Lanata, A., & Scilingo, E. P. (2015). Recognizing emotions induced by affective sounds through heart rate variability. *IEEE Transactions on Affective Computing*, *6*(4), 385–394.

Olbrich, S., Tränkner, A., Surova, G., Gevirtz, R., Gordon, E., Hegerl, U., & Arns, M. (2016). Cns-and ans-arousal predict response to antidepressant medication: Findings from the randomized ispot-d study. *Journal of Psychiatric Research*, *73*, 108e115.

pandas development team, T. (2020). *Pandas-dev/pandas: Pandas* (Version latest). Zenodo. https://doi.org/10.5281/zenodo.3509134

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

Pernet, C. R., Appelhoff, S., Gorgolewski, K. J., Flandin, G., Phillips, C., Delorme, A., & Oostenveld, R. (2019). Eeg-bids, an extension to the brain imaging data structure for electroencephalography. *Scientific data*, *6*(1), 1–5.

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology*, *17*(3), 715–734.

Ramoser, H., Muller-Gerking, J., & Pfurtscheller, G. (2000). Optimal spatial filtering of single trial eeg during imagined hand movement. *IEEE transactions on rehabilitation engineering*, *8*(4), 441–446.

Ringle, C. M., Wende, S., & Becker, J.-M. (2015). Smartpls 3. bönningstedt: Smartpls.

Roy, J.-C., Sequeira, H., & Delerm, B. (1993). Neural control of electrodermal activity: Spinal and reticular mechanisms. *Progress in electrodermal research* (pp. 73–92). Springer.

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

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of personality and social psychology*, *76*(5), 805.

Sabbagh, D., Ablin, P., Varoquaux, G., Gramfort, A., & Engemann, D. A. (2020). Predictive regression modeling with meg/eeg: From source power to signals and cognitive states. *NeuroImage*, *222*, 116893.

Sato, W., Kochiyama, T., & Yoshikawa, S. (2020). Physiological correlates of subjective emotional valence and arousal dynamics while viewing films. *Biological Psychology*, *157*, 107974.

Schell, M. E. D. A. M., & Filion, D. L. (2007). The electrodermal system. *Principles of Psychophysiology: Physical, social and inferential elements*, 295–324.

Schiappa, C., Scarpelli, S., D'Atri, A., Gorgoni, M., & De Gennaro, L. (2018). Narcolepsy and emotional experience: A review of the literature. *Behavioral and Brain functions*, *14*(1), 1–11.

Schlosberg, H. (1952). The description of facial expressions in terms of two dimensions. *Journal of experimental psychology*, *44*(4), 229.

Schmidt, F. M., Sander, C., Dietz, M.-E., Nowak, C., Schröder, T., Mergl, R., Schönknecht, P., Himmerich, H., & Hegerl, U. (2017). Brain arousal regulation as response predictor for antidepressant therapy in major depression. *Scientific reports*, *7*(1), 1–10.

Schoffelen, J.-M., Poort, J., Oostenveld, R., & Fries, P. (2011). Selective movement preparation is subserved by selective increases in corticomuscular gamma-band coherence. *Journal of Neuroscience*, *31*(18), 6750–6758.

Sequeira, H., & Roy, J.-C. (1993). Cortical and hypothalamo-limbic control of electrodermal responses. *Progress in electrodermal research* (pp. 93–114). Springer.

Shukla, J., Barreda-Angeles, M., Oliver, J., Nandi, G., & Puig, D. (2019). Feature extraction and selection for emotion recognition from electrodermal activity. *IEEE Transactions on Affective Computing*.

Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-n design. *Psychonomic bulletin & review*, *25*(6), 2083–2101.

Snijders, T. A. (2005). Power and sample size in multilevel modeling. *Encyclopedia of statistics in behavioral science*, *3*(157), 1573.

Soleymani, M., Lichtenauer, J., Pun, T., & Pantic, M. (2011). A multimodal database for affect recognition and implicit tagging. *IEEE transactions on affective computing*, *3*(1), 42–55.

Storbeck, J., & Clore, G. L. (2008). Affective arousal as information: How affective arousal influences judgments, learning, and memory. *Social and personality psychology compass*, *2*(5), 1824–1843.

Uusitalo, M. A., & Ilmoniemi, R. J. (1997). Signal-space projection method for separating meg or eeg into components. *Medical and Biological Engineering and Computing*, *35*(2), 135–140.

Van Boxtel, A. (2001). Optimal signal bandwidth for the recording of surface emg activity of facial, jaw, oral, and neck muscles. *Psychophysiology*, *38*(1), 22–34.

Van den Burg, W., Beersma, D., Bouhuys, A., & Van den Hoofdakker, R. (1992). Self-rated arousal concurrent with the antidepressant response to total sleep deprivation of patients with a major depressive disorder: A disinhibition hypothesis. *Journal of Sleep Research*, *1*(4), 211–222.

Vellido, A. (2020). The importance of interpretability and visualization in machine learning for applications in medicine and health care. *Neural Computing and Applications*, *32*(24), 18069–18083.

Wilson-Mendenhall, C. D., Barrett, L. F., & Barsalou, L. W. (2013). Neural evidence that human emotions share core affective properties. *Psychological science*, *24*(6), 947–956.

Wundt, W. M. (1912). *An introduction to psychology*. G. Allen, Limited.

Yik, M. S., Russell, J. A., & Barrett, L. F. (1999). Structure of self-reported current affect: Integration and beyond. *Journal of personality and social psychology*, *77*(3), 600.

Ziemssen, T., & Siepmann, T. (2019). The investigation of the cardiovascular and sudomotor autonomic nervous system—a review. *Frontiers in neurology*, *10*, 53.

7 Supplementary Materials

7.1 Figures

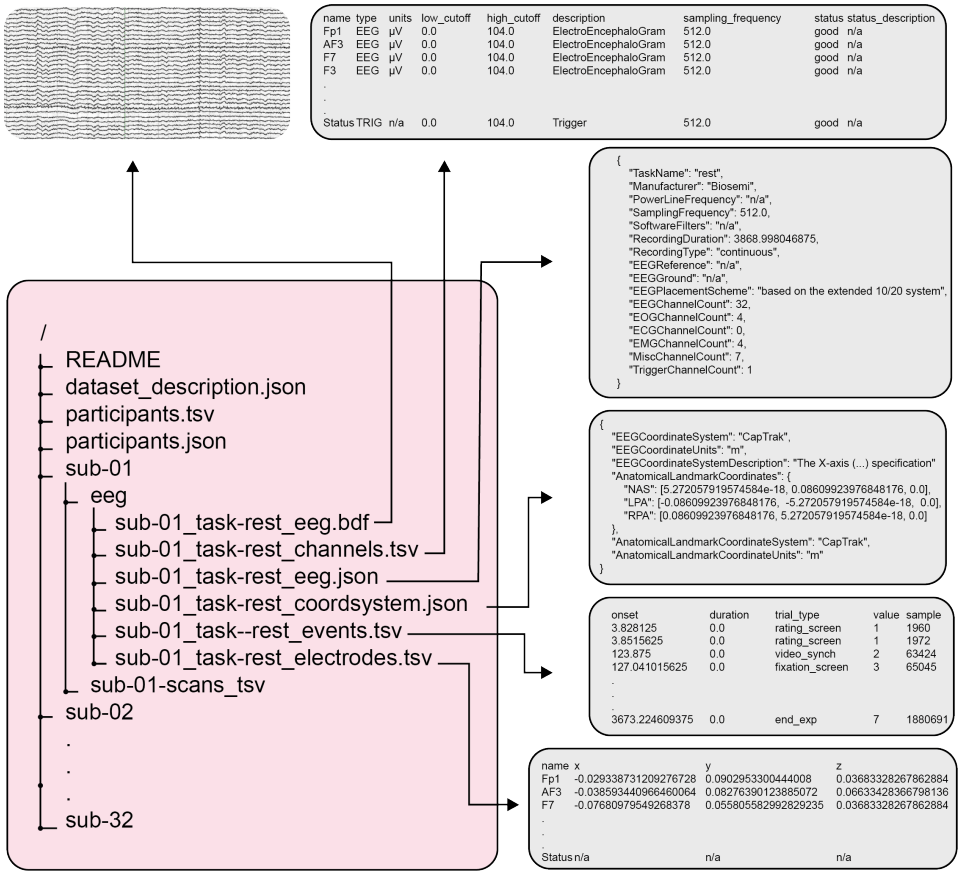

Figure S1. DEAP BIDS organization (with focus on Subject 01). This representation includes not only the raw data (*sub-01_task-rest_eeg.bdf*), but also additional information on channels (*sub-01_task-rest_channels.tsv*) and the general recording (*sub-01_task-rest_eeg.json*), information regarding electrode locations (*sub-01_task-rest_coordsystem.json* and *sub-01_task-rest_electrodes.tsv*) and also about recording events (*sub-01_task-rest_events.tsv*).

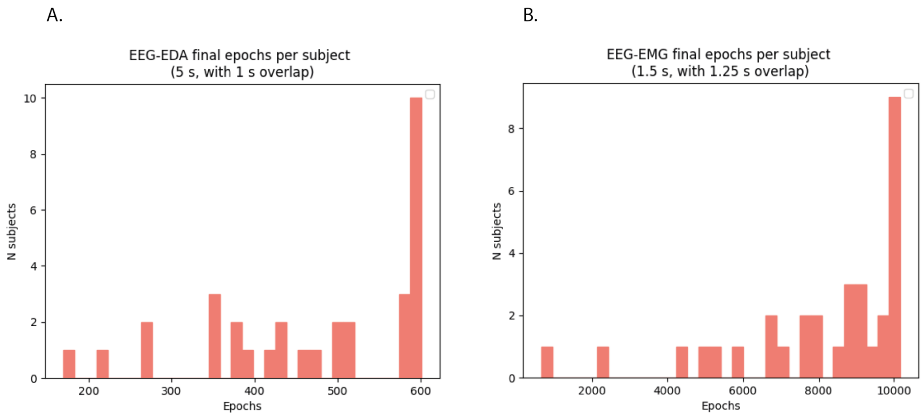

Figure S2. Distribution of the final number of epochs per subject for both **A**. EEG-EDA and **B.** EEG-EMG epochs.

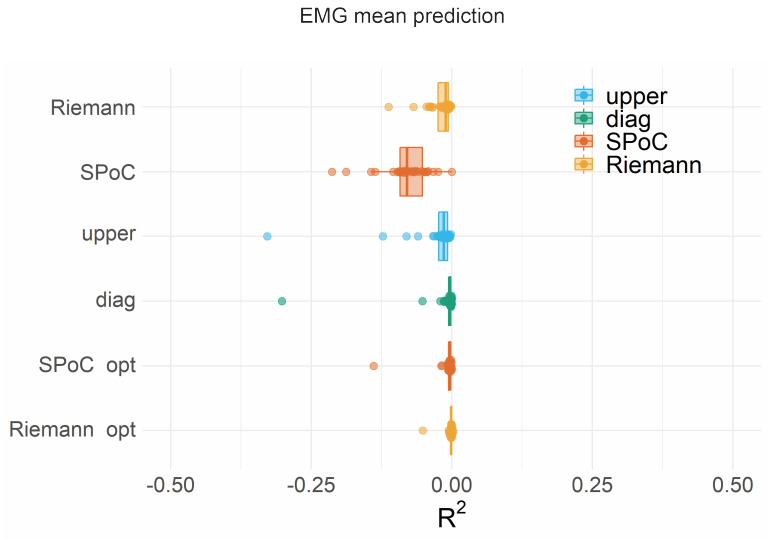

Figure S3. Model comparison in EMG mean prediction across subjects. The different models to be compared are represented on the Y-axis of each subfigure, while performance is represented on the X-axis. Each type of model has a different colour in the figure. Thus, dots represent the average expected out of sample performance (R^2) of the 2-fold cross validation of each participant. The distribution of participants' performance on each model is summarized by boxplots.

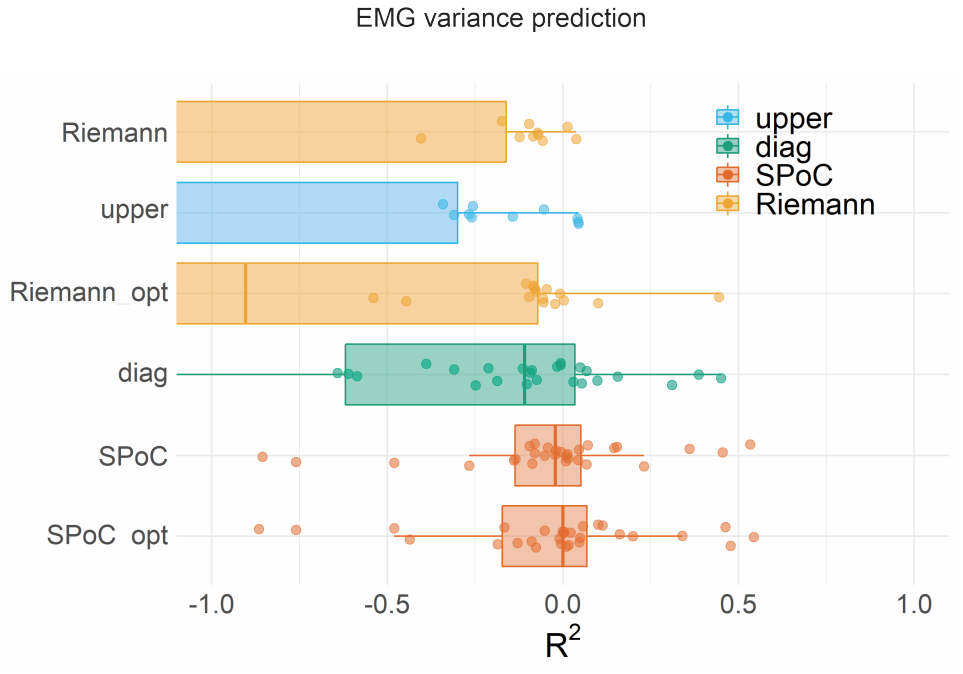

Figure S3. Model comparison in EMG variance prediction across subjects. The different models to be compared are represented on the Y-axis of each subfigure, while performance is represented on the X-axis. Each type of model has a different colour in the figure. Thus, dots represent the average expected out of sample performance (R^2) of the 2-fold cross validation of each participant. The distribution of participants' performance on each model is summarized by boxplots.

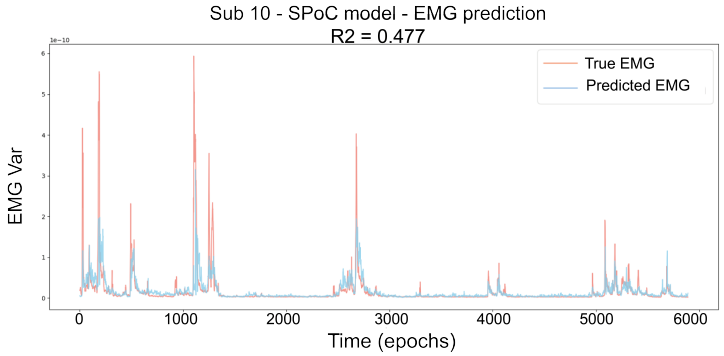

Figure S4. Continuous EMG variance decoding with *SPoC* model for subject 10. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EMG variance values. The true signal (i.e. EMG variance) is represented in red, while the predicted signal (i.e. predicted EMG variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

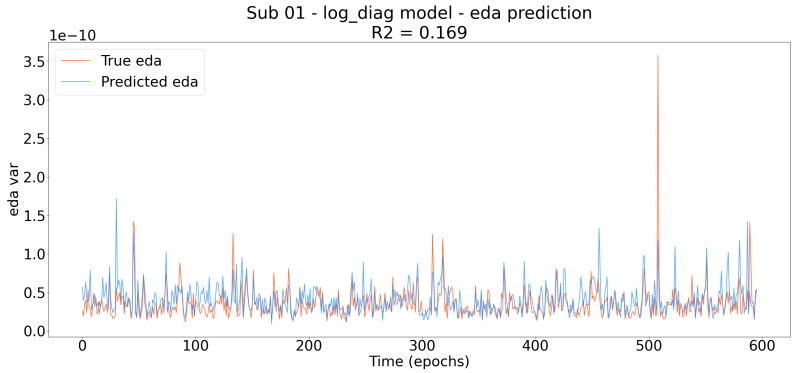

Figure S5. Continuous EDA variance decoding with *diag* model for subject 01. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

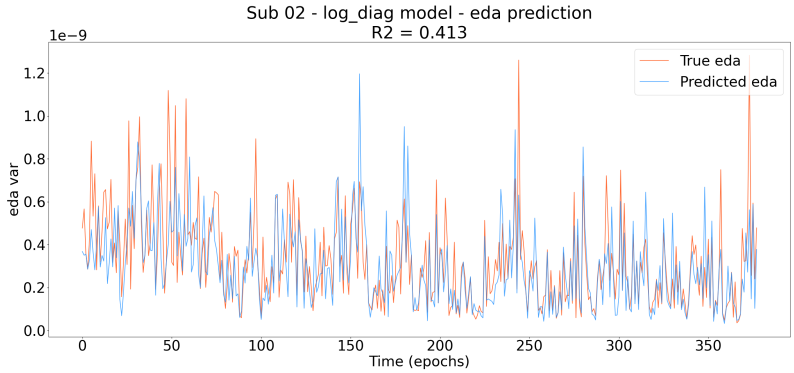

Figure S5. Continuous EDA variance decoding with *diag* model for subject 02. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

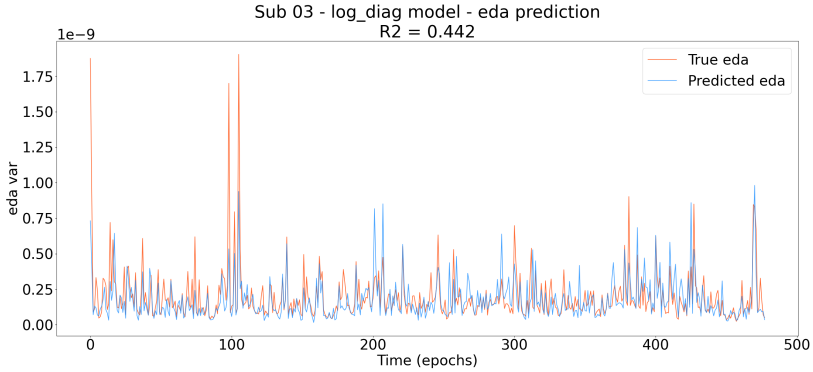

Figure S5. Continuous EDA variance decoding with *diag* model for subject 03. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

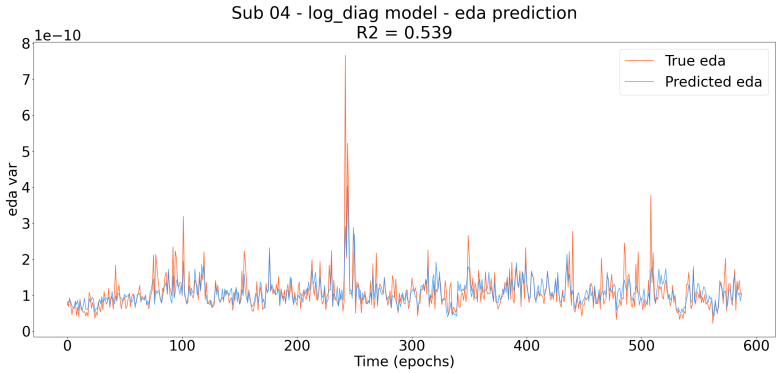

Figure S5. Continuous EDA variance decoding with *diag* model for subject 04. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

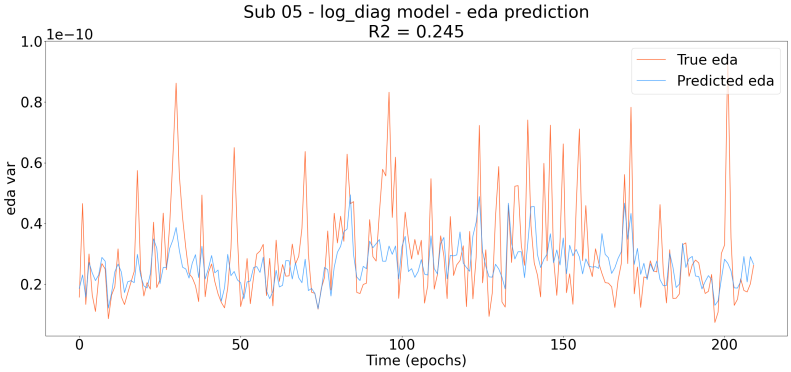

Figure S5. Continuous EDA variance decoding with *diag* model for subject 05. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

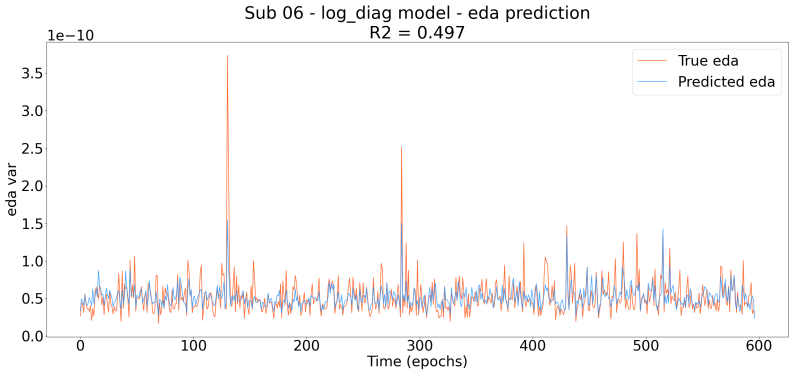

Figure S5. Continuous EDA variance decoding with *diag* model for subject 06. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

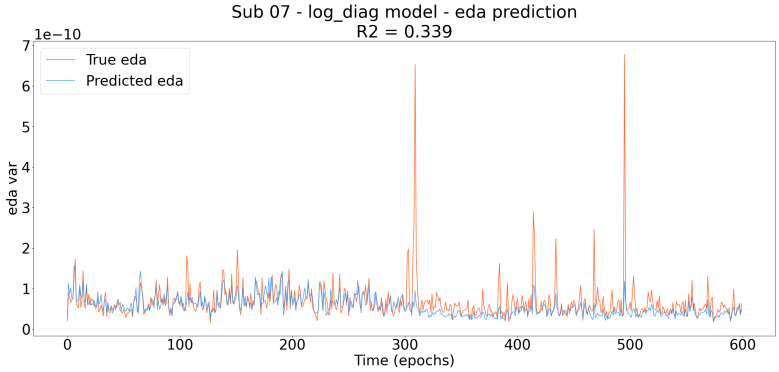

Figure S5. Continuous EDA variance decoding with *diag* model for subject 07. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

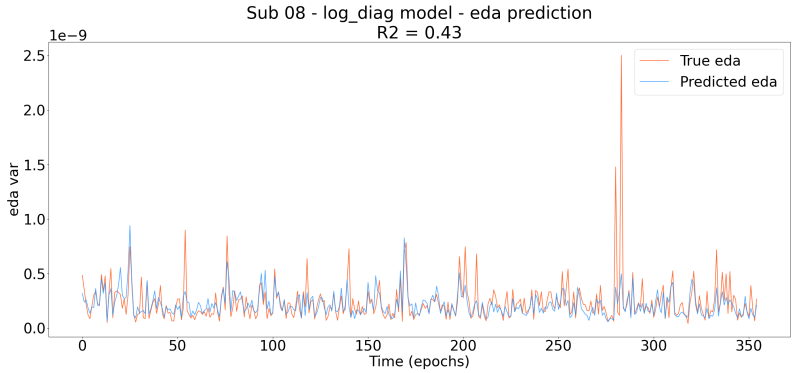

Figure S5. Continuous EDA variance decoding with *diag* model for subject 08. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

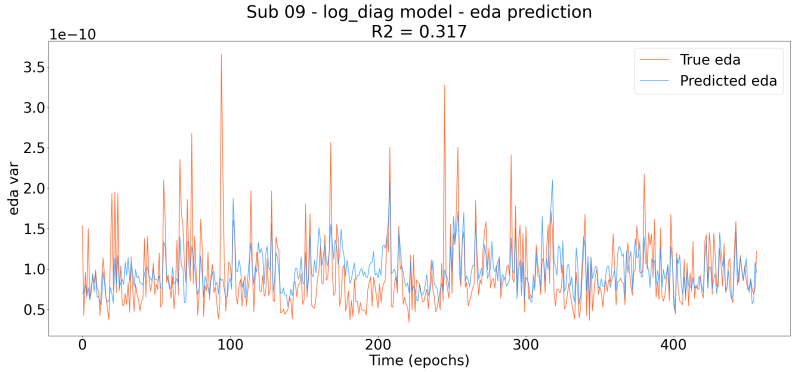

Figure S5. Continuous EDA variance decoding with *diag* model for subject 09. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

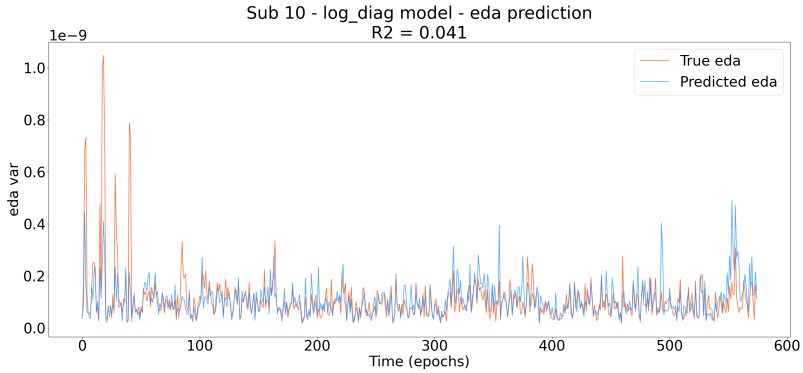

Figure S5. Continuous EDA variance decoding with *diag* model for subject 10. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

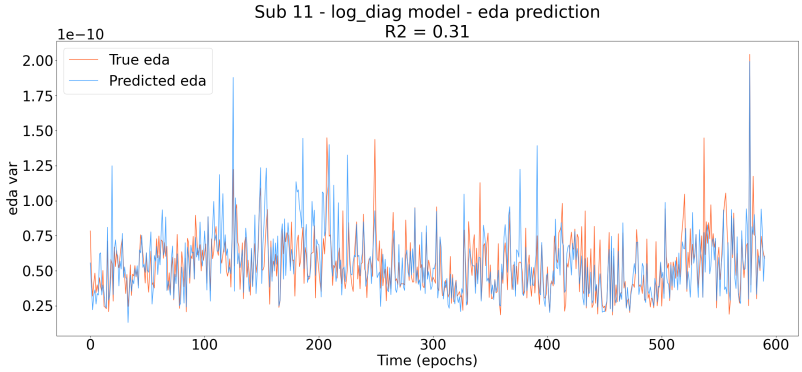

Figure S5. Continuous EDA variance decoding with *diag* model for subject 11. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

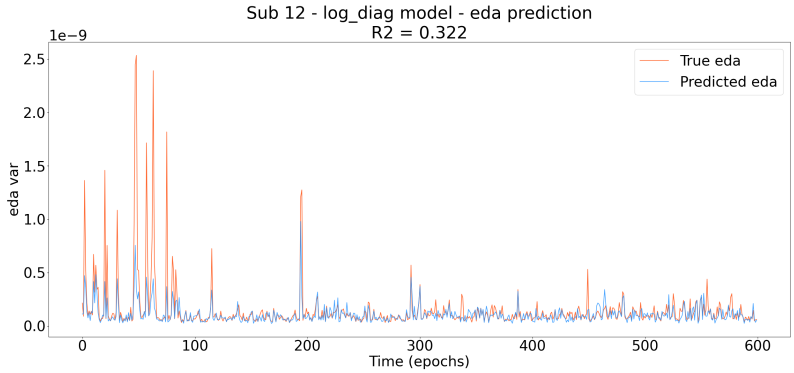

Figure S5. Continuous EDA variance decoding with *diag* model for subject 12. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

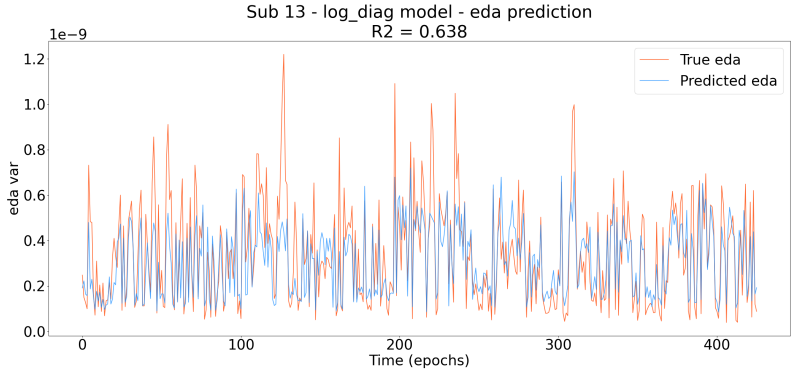

Figure S5. Continuous EDA variance decoding with *diag* model for subject 13. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

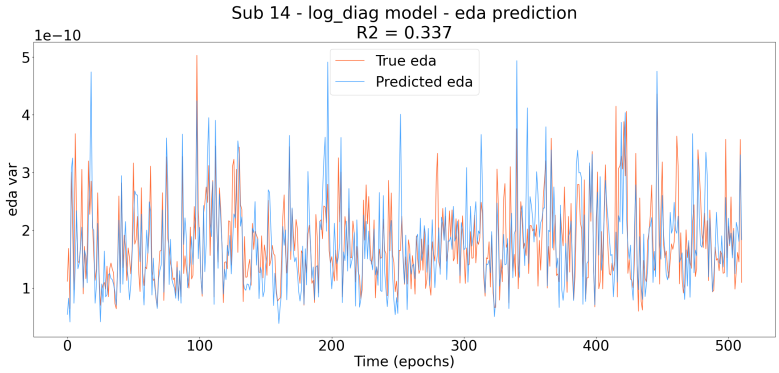

Figure S5. Continuous EDA variance decoding with *diag* model for subject 14. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

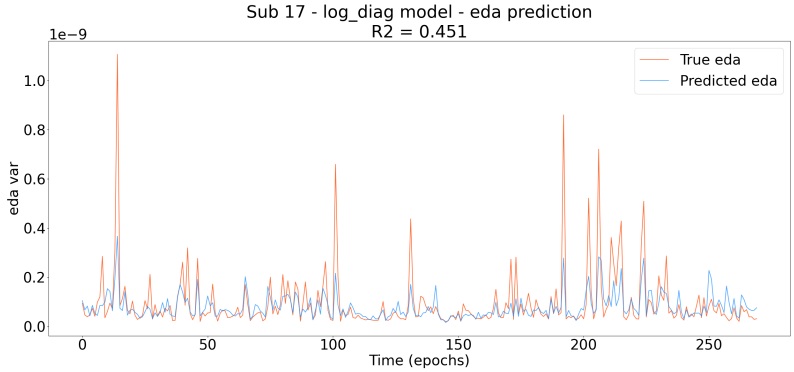

Figure S5. Continuous EDA variance decoding with *diag* model for subject 17. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

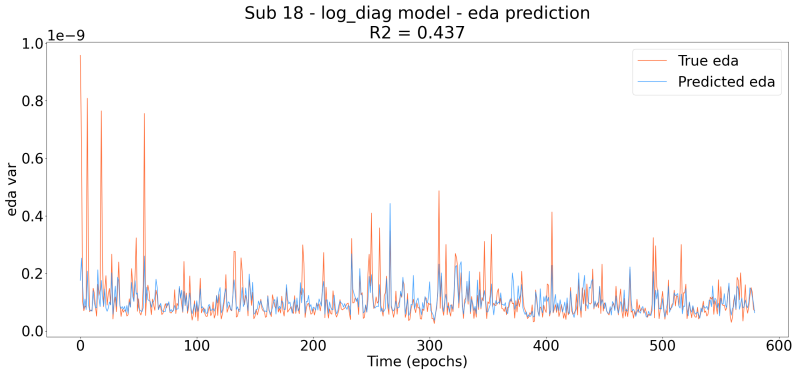

Figure S5. Continuous EDA variance decoding with *diag* model for subject 18. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

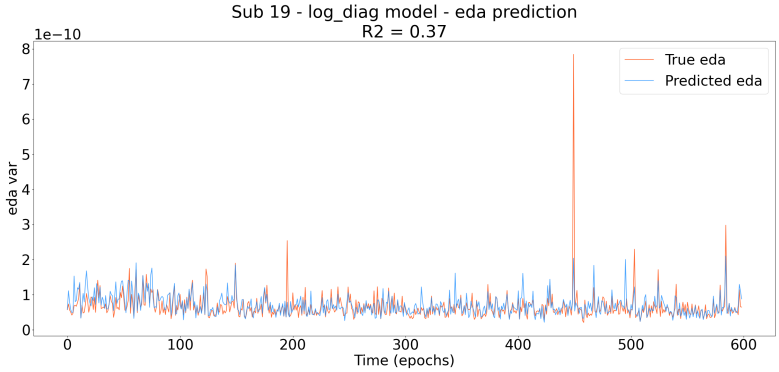

Figure S5. Continuous EDA variance decoding with *diag* model for subject 19. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

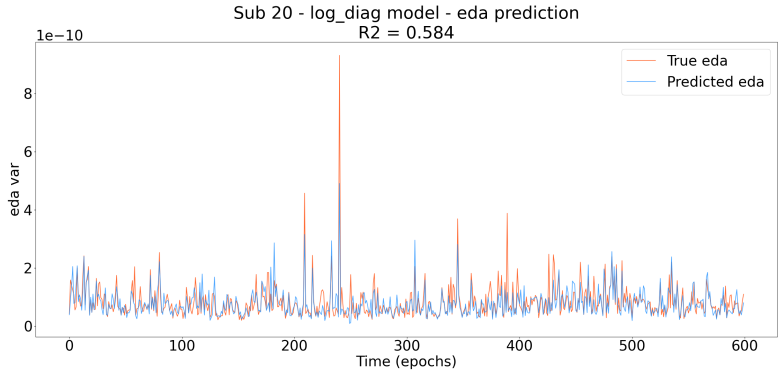

Figure S5. Continuous EDA variance decoding with *diag* model for subject 20. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

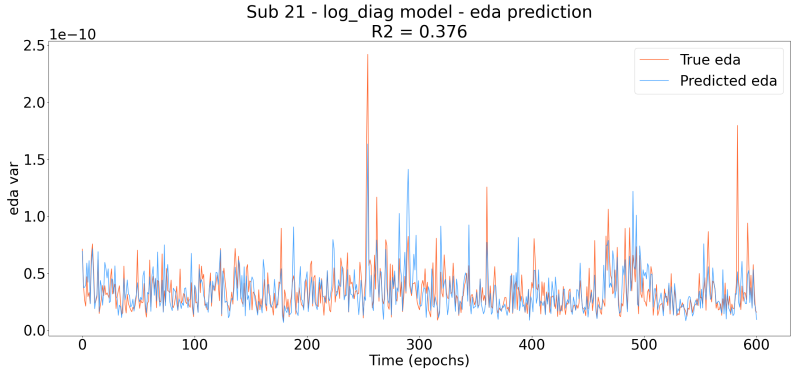

Figure S5. Continuous EDA variance decoding with *diag* model for subject 21. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

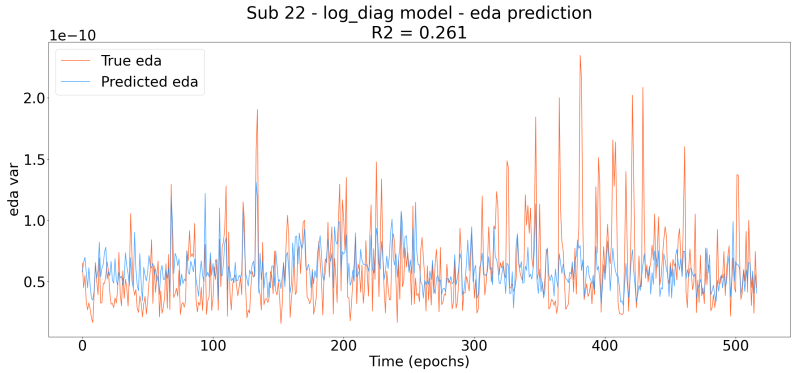

Figure S5. Continuous EDA variance decoding with *diag* model for subject 22. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

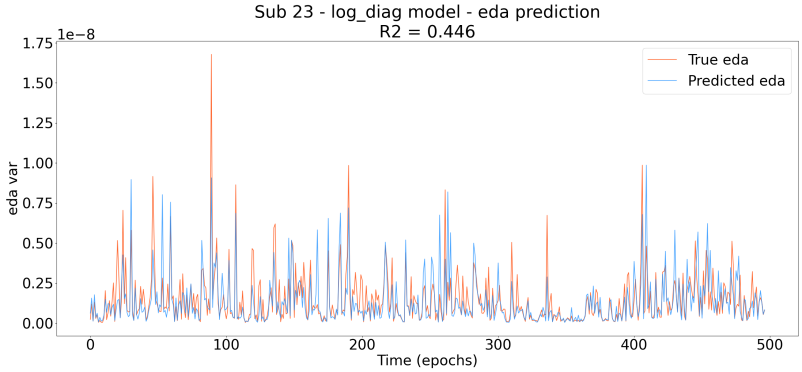

Figure S5. Continuous EDA variance decoding with *diag* model for subject 23. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

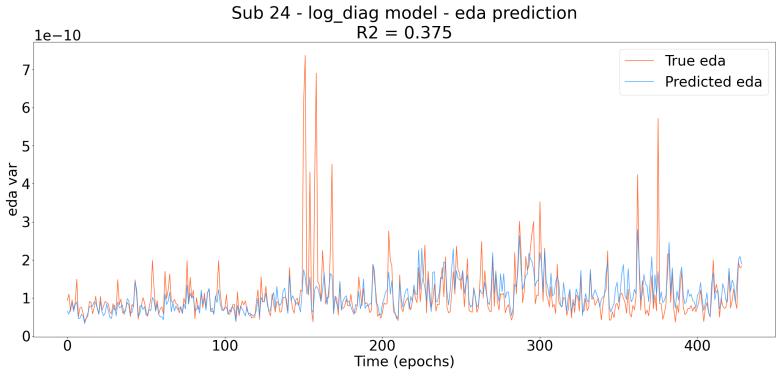

Figure S5. Continuous EDA variance decoding with *diag* model for subject 24. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

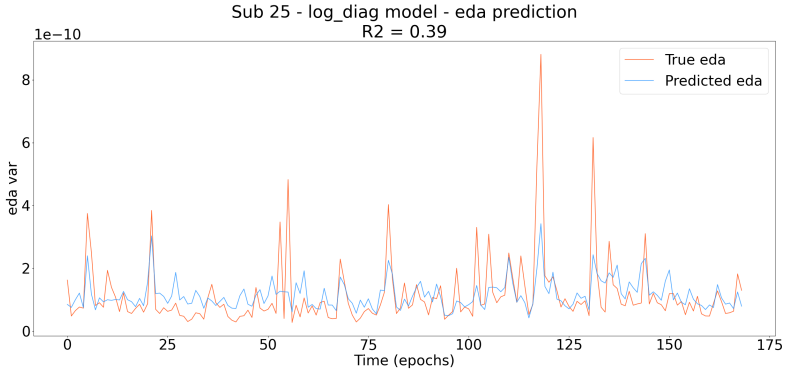

Figure S5. Continuous EDA variance decoding with *diag* model for subject 25. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

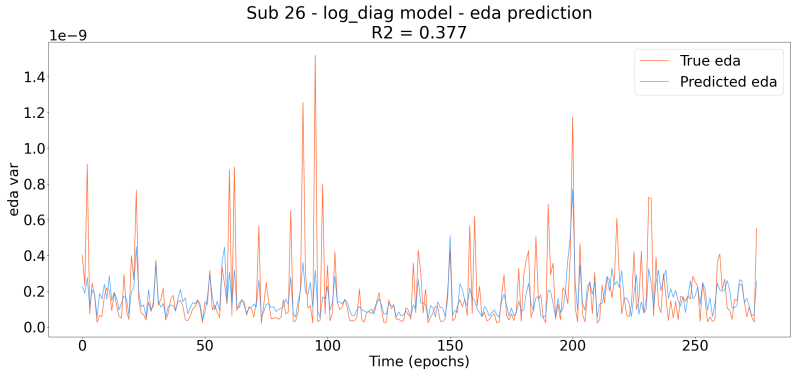

Figure S5. Continuous EDA variance decoding with *diag* model for subject 26. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

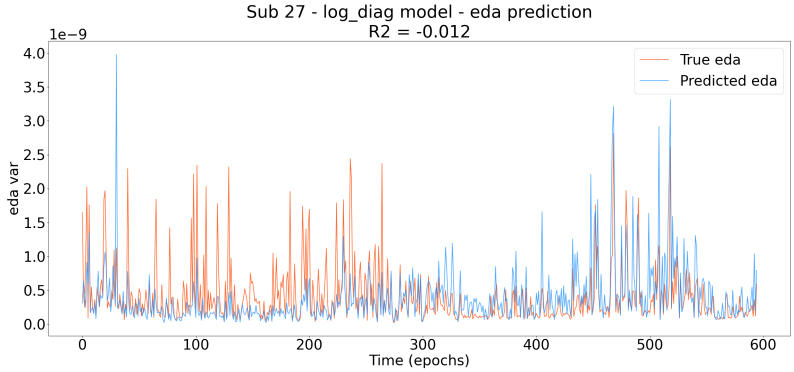

Figure S5. Continuous EDA variance decoding with *diag* model for subject 27. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

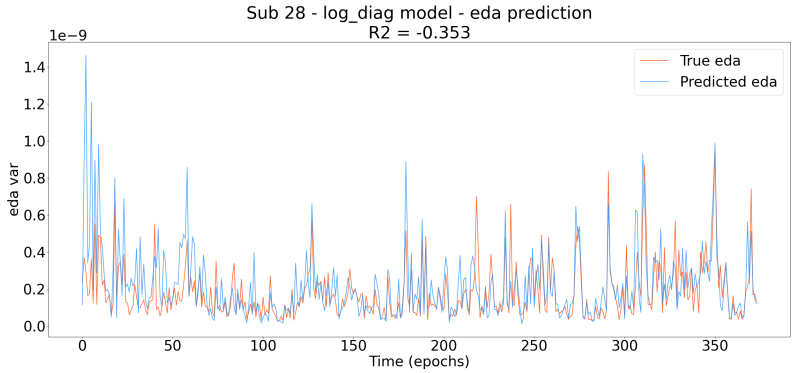

Figure S5. Continuous EDA variance decoding with *diag* model for subject 28. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

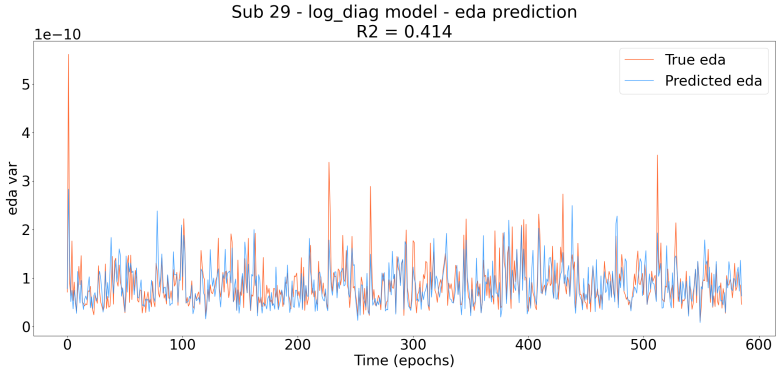

Figure S5. Continuous EDA variance decoding with *diag* model for subject 29. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

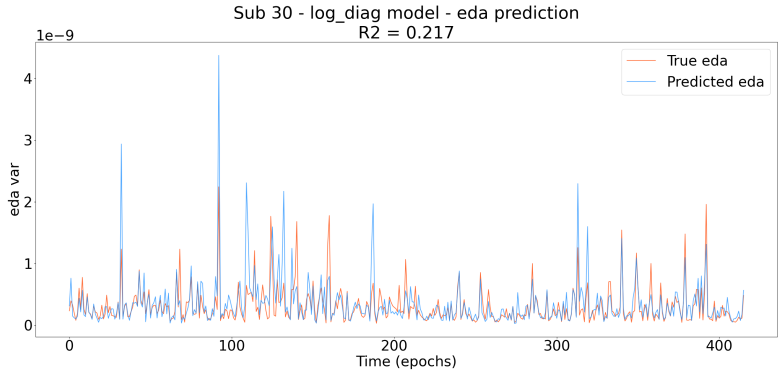

Figure S5. Continuous EDA variance decoding with *diag* model for subject 30. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

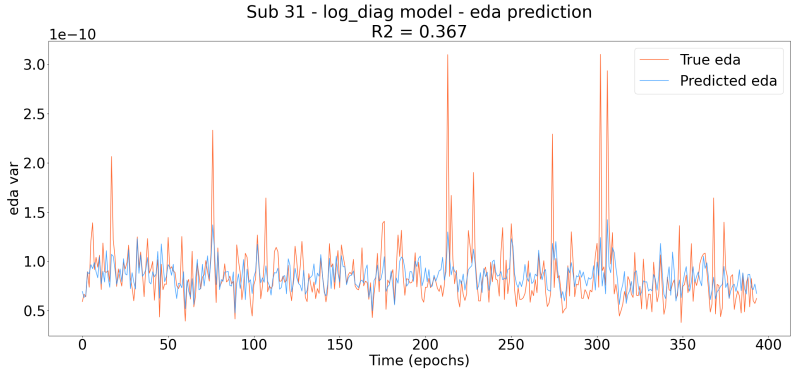

Figure S5. Continuous EDA variance decoding with *diag* model for subject 31. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

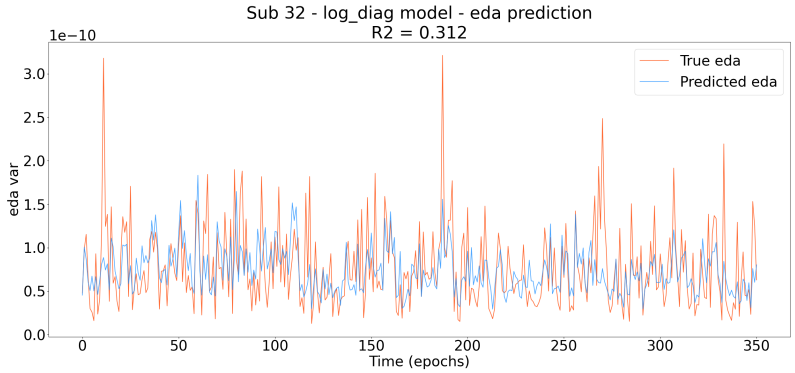

Figure S5. Continuous EDA variance decoding with *diag* model for subject 32. The x-axis of each figure represents the time (in epochs), while the y-axis shows the EDA variance values. The true signal (i.e. EDA variance) is represented in red, while the predicted signal (i.e. predicted EDA variance) is represented in blue. It is specified the R^2 obtained in this subject's prediction.

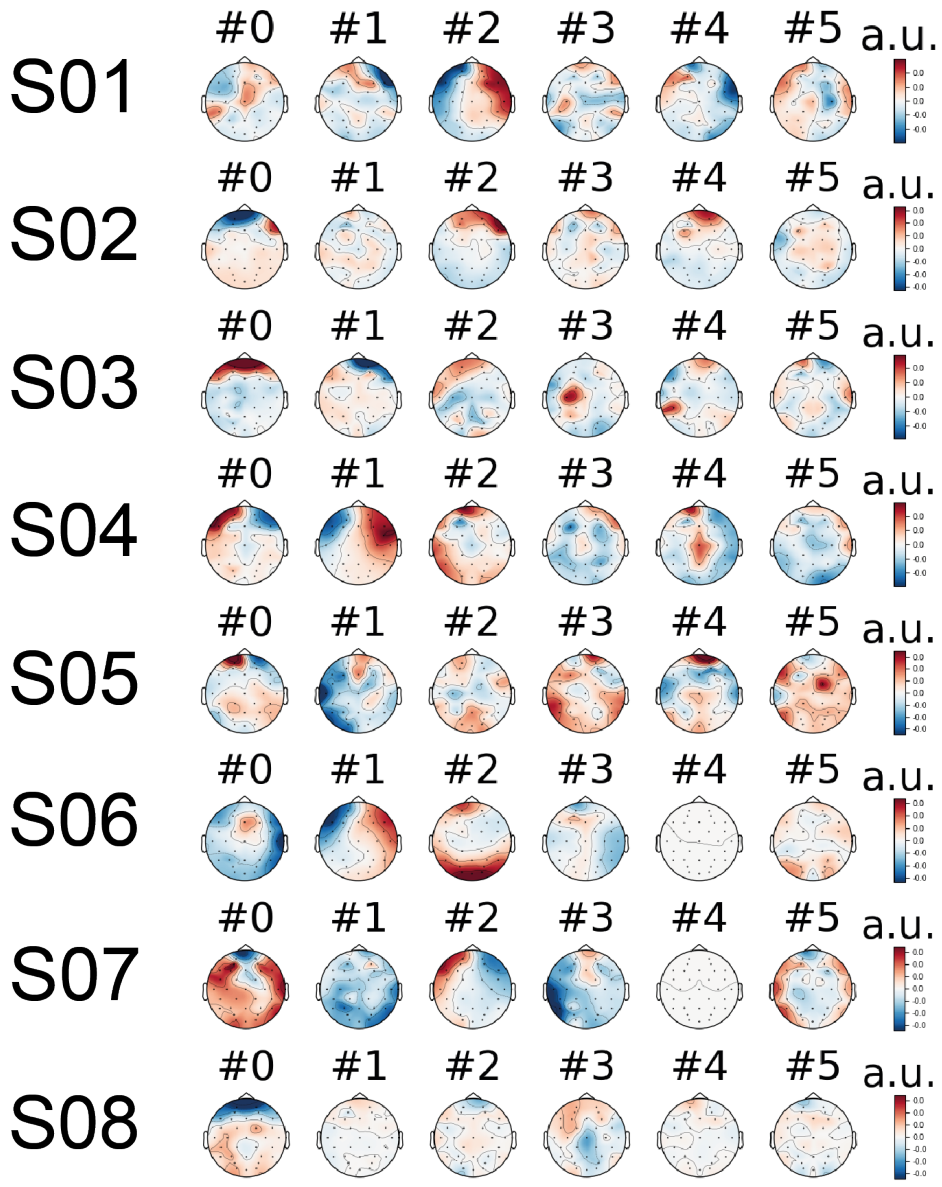

Figure S6. Plot of the contribution of the six first detected components in EDA variance decoding task for subjects 01 to 08. Colours indicate the contribution of that sensor space to the SPoC patterns represented.

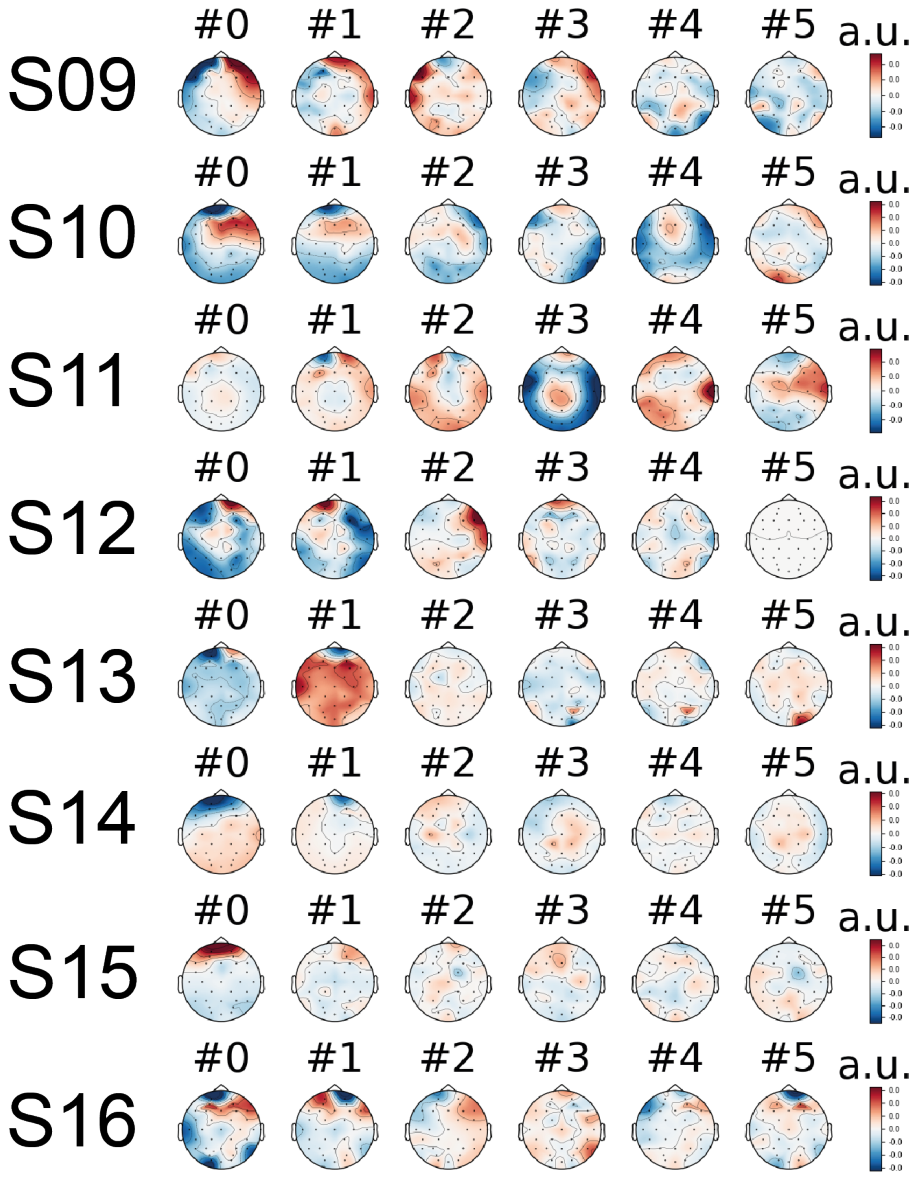

Figure S6. Plot of the contribution of the six first detected components in EDA variance decoding task for subjects 09 to 16. Colours indicate the contribution of that sensor space to the SPoC patterns represented.

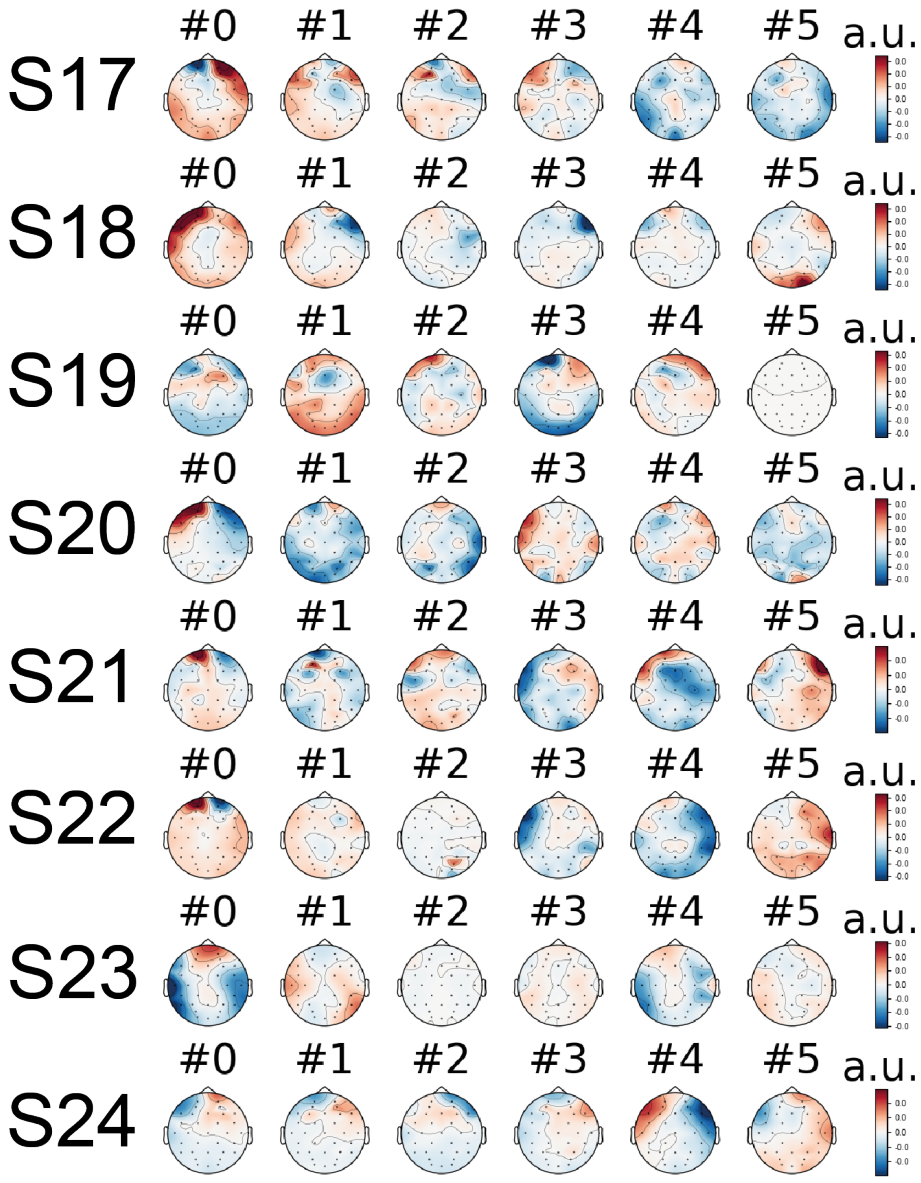

Figure S6. Plot of the contribution of the six first detected components in EDA variance decoding task for subjects 17 to 24. Colours indicate the contribution of that sensor space to the SPoC patterns represented.

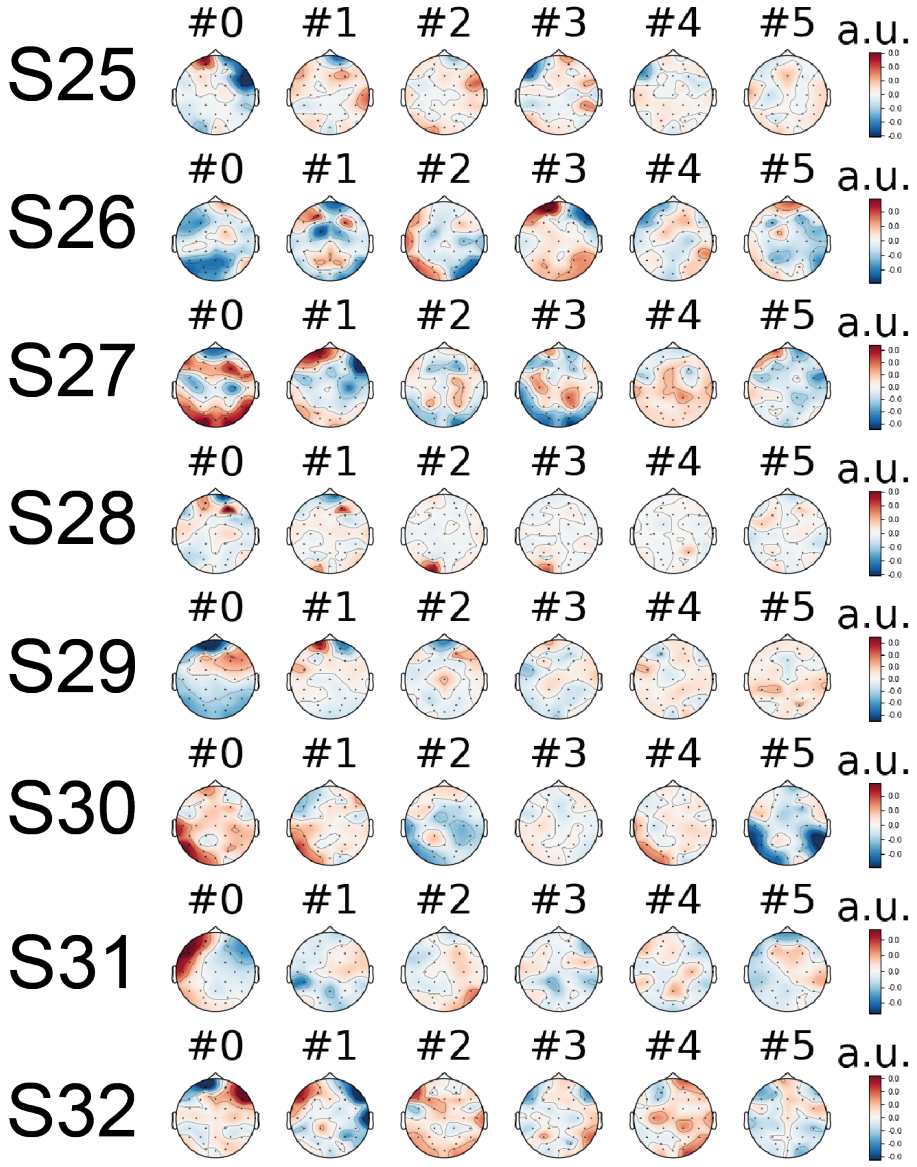

Figure S6. Plot of the contribution of the six first detected components in EDA variance decoding task for subjects 25 to 32. Colours indicate the contribution of that sensor space to the SPoC patterns represented.

7.2 Tables

Status code (subject 1 to 23)	Status code (subject 24 to 32)	Event duration	Event Description
1	1638145	N/A	First occurence: start of experiment (participant pressed key to start)
1	1638145	120000 ms	Second occurence: start of baseline recording
1	1638145	N/A	Further occurences: start of a rating screen
2	1638146	1000 ms	Video synchronization screen (before first trial, before and after break, after last trial)
3	1638147	5000 ms	Fixation screen before beginning of trial
4	1638148	60000 ms	Start of music video playback
5	1638149	3000 ms	Fixation screen after music video playback
7	1638151	N/A	End of experiment

Table S1. Status codes and descriptions in DEAP database from subjects 1 to 23 (https://www.eecs.qmul.ac.uk/mmv/datasets/deap/readme.html); and inferred status codes and descriptions from subjctes 24 to 32 according to visual inspection of events.

8 Preregistration Document

Tomás A. D'Amelio
Supervisor: Denis A. Engemann

CogMaster Preregistration

0. Administrative information

- **Member of the conseil pédagogique**: Claire Sergent
- **External researcher**: Vadim Nikulin or Lucas Parra or Enzo Tagliazucchi

1. Introduction

Background and rationale. In the last years, affective arousal prediction has received increasing attention in engineering and computer science literature. Among other biosignals, electrodermal activity (EDA) and EEG are used to recognize participants' self-reported affective states during an emotion elicitation task.

The majority of these studies focus on event-level (e.g emotional response to a picture) statistical learning models (Sabbagh et al., 2020), as it is difficult to achieve a sufficient sample size for modelling subject-level outcomes (e.g. mood, diagnosis). In the last years, machine learning at the event-level has been used to enhance subject-level analysis (King et al., 2013; Sitt et al., 2014). What should be done if there are multiple levels of events at different time-scales? This is the case in arousal decoding where the EDA tracks arousal continuously and subjective judgments are recorded once per trial.

At the physiological level, it is assumed that the information provided by EEG and EDA signals is closely related to the generation of arousal. However, both signals can provide different information to describe subjective arousal by complementing peripheral information (i.e. EDA) with data related to the central nervous system (i.e. EEG). Can we enhance the analysis of arousal self-reports by summarizing the relationship between EDA and EEG using machine learning?

Recent work in population modeling with neuroscience data suggest that an outcome predicted from brain signals using machine learning can enrich the measured outcome of interest, hence, yielding a proxy measure (Engemann et al., 2020; Dadi et al., 2020). Can this approach be translated to the problem of arousal decoding?

To investigate these questions, we will analyze here a public dataset containing both physiological responses (i.e. EEG and EDA) and self-reported arousal through exposure to emotional stimuli. The task to be learned by the model will be self-reported arousal regression. However, instead of directly predicting this measure from the available physiological signals, a two-step procedure is proposed: first, electrodermal activity will be predicted from EEG; then, self-reported arousal will be decoded from the regressed EDA.

Key research question. Can the capacity of EDA to capture arousal be enhanced by approximating EDA from EEG?

General hypotheses. It is hypothesized that models trained with predicted EDA will enhance modeling of subjective arousal beyond the observed EDA.

2. Methods

Model

Here we propose a two-step approach based on general purpose statistical methods (i.e. the models will be unrelated to the task). First, EDA decoding from EEG will be performed, with the idea of boosting high-resolution signals in each subject. Thus, models with continuous inputs (i.e. EDA) and outputs (i.e. EEG) will be trained, in order to extract as much information as possible from each subject to learn different subject-level function approximations. As an output of this first step, arousal will be represented with the predicted EDA. This predicted signal is expected to capture the coupling of autonomic and cerebral arousal, as it is the result of the product of the EEG features weighted by the different coefficients generated after the fitting process with the EDA output data. This would allow to make use of the richness of these biosignals in a unique representation at the subject level. In this way, the predicted arousal would contain information that is not included in the original EDA data. Consequently, in a second step, a statistical learning model to predict self-reported arousal from the predicted EDA will be trained.

Input data / material

The Database for Emotion Analysis using Physiological Signals (DEAP; Koelstra et al., 2012) will be analyzed, which contains both EEG and EDA participant recordings during an emotional elicitation task will be analyzed. DEAP contains information of 32 subjects (16 females) who were recorded as each watched 40 musical emotional stimuli of one-minute duration (with 3 seconds of pre-trial recording). In this database, besides valence and arousal, there are also dominance, liking, and familiarity scores that correspond with each participant's trial. In addition to EEG and EDA, DEAP has registered horizontal and vertical EOG, zygomaticus major and trapezius.

A minimal preprocessing to both raw EEG and EDA data will be applied: recordings would be downsampled to 250Hz and data segments dominated by high-amplitude signals would be excluded signals using the 'global' option from autoreject (Jas et al., 2017).

Measures (output)

The output measure will be subjective arousal value, with a range between 1 to 9 (from low to high self reported arousal, respectively)

Predictions

We expect that both predicted and observed EDA will show additive contributions (significant coefficients) in a linear mixed-effects model with stimulus and subject as random effects.

Analyses

EDA modeling Modelling will be conducted using Python. For processing of EEG we will use the MNE package (Gramfort et al., 2014). For machine learning, we will use the Scikit-Learn library (Pedregosa et al., 2011). We will approximate EDA from EEG as a weighted some of spatial information in different frequency bands using a linear model. As EDA is a positive continuous outcome, we will use a Gamma likelihood. To correct for distortions due to field spread we will use the Reimannian embedding of the sensor-space covariance in filter-bank style as in Sabbagh et al. (2019, 2020).

We will therefore constrain parameter fitting with an L2 (ridge) penalty that will push the solution towards the principal components that explain more variance. Hyper-parameter search will be controlled through grid search (nested cross-validation). Overfitting will be avoided by generating the predicted EDA using 2-fold cross-validation (adapted for time-series to avoid auto-correlation). Uncertainty estimates will be obtained from non-parametric bootstrapping of the entire process with 2000 iterations.

Self-report and subject-level modeling. The predicted EDA, alongside the observed EDA will then be subjected to a linear mixed-effects model with stimulus and subject as random effects.

Interpretation

Our proposed model will have "observed EDA" and "EEG-enriched EDA" as independent variables for the subjective arousal regression analysis. Thus, if we find an additive contribution of the predicted EDA to the observed EDA, this would mean that the predicted EDA would have information that is not redundant for the arousal regression task compared to the observed EDA.

3. Expected contributions

Finding this result would imply that it is possible to trace the generating sources of EDA through EEG (controlling by error sources; e.g. heartbeats or eye blinks). In this way, it would also be possible to infer which are the most relevant EEG markers (i.e. features) for EDA prediction, an issue that has not been done previously in the literature, according to the review performed by the authors.

4. References

Dadi, K., Varoquaux, G., Houenou, J., Bzdok, D., Thirion, B., & Engemann, D. (2020). *Beyond brain age: Empirically-derived proxy measures of mental health*. https://doi.org/10.1101/2020.08.25.266536

Engemann, D. A., Kozynets, O., Sabbagh, D., Lemaître, G., Varoquaux, G., Liem, F., & Gramfort, A. (2020). Combining magnetoencephalography with magnetic resonance imaging enhances learning of surrogate-biomarkers. *ELife*, *9*, e54055. https://doi.org/10.7554/eLife.54055

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Parkkonen, L., & Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage*, *86*, 446–460. https://doi.org/10.1016/j.neuroimage.2013.10.027

Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F., & Gramfort, A. (2017). Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage*, *159*, 417–429. https://doi.org/10.1016/j.neuroimage.2017.06.030

King, J. R., Faugeras, F., Gramfort, A., Schurger, A., El Karoui, I., Sitt, J. D., Rohaut, B., Wacongne, C., Labyt, E., Bekinschtein, T., Cohen, L., Naccache, L., & Dehaene, S. (2013). Single-trial decoding of auditory novelty responses facilitates the detection of residual consciousness. *NeuroImage*, *83*, 726–738. https://doi.org/10.1016/j.neuroimage.2013.07.013

Koelstra, S., Muhl, C., Soleymani, M., Lee, J., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., & Patras, I. (2012). DEAP: A Database for Emotion Analysis ;Using Physiological Signals. *IEEE Transactions on Affective Computing*, *3*(1), 18–31. https://doi.org/10.1109/T-AFFC.2011.15

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. https://hal.inria.fr/hal-00650905

Sabbagh, D., Ablin, P., Varoquaux, G., Gramfort, A., & Engemann, D. A. (2019). Manifold-regression to predict from MEG/EEG brain signals without source modeling. *Advances in Neural Information Processing Systems*, *32*, 7323–7334.

Sabbagh, D., Ablin, P., Varoquaux, G., Gramfort, A., & Engemann, D. A. (2020). Predictive regression modeling with MEG/EEG: From source power to signals and cognitive states. *NeuroImage*, *222*, 116893. https://doi.org/10.1016/j.neuroimage.2020.116893

Sitt, J. D., King, J.-R., El Karoui, I., Rohaut, B., Faugeras, F., Gramfort, A., Cohen, L., Sigman, M., Dehaene, S., & Naccache, L. (2014). Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. *Brain*, *137*(8), 2258–2270. https://doi.org/10.1093/brain/awu141