Decoding subjective emotional arousal from EEG during an immersive Virtual Reality experience

Simon M. Hofmann[1*], Felix Klotzsche[1,4*], Alberto Mariola[2,3*], Vadim V. Nikulin[1,4], Arno Villringer[1,4], Michael Gaebler[1,4]

[1] Department of Neurology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

[2] Sussex Neuroscience, University of Sussex, Brighton, UK

[3] Sackler Centre for Consciousness Science, University of Sussex, Brighton, UK

[4] MindBrainBody Institute at the Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Berlin, Germany

* shared authorship, alphabetical order

Correspondence

Simon M. Hofmann: simon.hofmann@cbs.mpg.de

Felix Klotzsche: klotzsche@cbs.mpg.de

Alberto Mariola: a.mariola@sussex.ac.uk

Michael Gaebler: gaebler@cbs.mpg.de

Key words

emotional arousal; EEG; machine learning; decoding; affective computing; deep learning; mind-brain-body; ecological validity; naturalistic stimuli; immersive VR; computational affective neuroscience

Abstract

Immersive virtual reality (VR) enables naturalistic neuroscientific studies while maintaining experimental control, but dynamic and interactive stimuli pose methodological challenges. We here probed the link between emotional arousal, a fundamental property of affective experience, and parieto-occipital alpha power under naturalistic stimulation:

37 young healthy adults completed an immersive VR experience, which included rollercoaster rides, while their EEG was recorded. They then continuously rated their subjective emotional arousal while viewing a replay of their experience. The association between emotional arousal and parieto-occipital alpha power was tested and confirmed by (1) decomposing the continuous EEG signal while maximizing the comodulation between alpha power and arousal ratings and by (2) decoding periods of high and low arousal with discriminative common spatial patterns and a Long Short-Term Memory recurrent neural network.

We successfully combine EEG and a naturalistic immersive VR experience to extend previous findings on the neurophysiology of emotional arousal towards real-world neuroscience.

Introduction

While humans almost constantly interact with complex, dynamic environments, lab-based studies typically use simplified stimuli in passive experimental situations. Trading realism for experimental control happens at the expense of the representativity of the experimental design (Brunswik, 1955), that is, the degree to which effects found in the lab generalize to practical everyday-life conditions. This is particularly true for affective phenomena like emotions.

Emotional arousal as a fundamental property of affective experience

Emotions are subjective, physiological, and behavioural responses to personally meaningful stimuli (Mauss & Robinson, 2009) that underlie our experience of the world (James, 1884, 1890; Seth, 2013). Emotions are crucial for physical and mental health (Gross & Muñoz, 1995) and their investigation has long been at the core of experimental psychology (Wundt, 1897).

Dimensional accounts conceptualize emotions along the two axes of valence and arousal (Duffy, 1957; Kuppens et al., 2013; Russell, 1980; Russell & Feldman Barrett, 1999; Wundt, 1897): Valence differentiates states of pleasure and displeasure, while (emotional) arousal describes the degree of activation or intensity that accompanies an emotional state.[1]

Emotions have been linked to activity in the autonomic (ANS) and the central nervous system (CNS; Dalgleish, 2004). It has thereby been difficult to consistently associate individual (i.e., discrete) emotion categories with specific response patterns in the ANS (cf. Kragel & LaBar, 2013; Kreibig, 2010; Siegel et al., 2018) or in distinct brain regions (Lindquist et al., 2012; but cf. Saarimäki et al., 2016). Rather, emotions seem to be dynamically implemented by sets of brain regions and bodily activations that are involved in basic psychological (i.e., also non-emotional) operations (i.e., "psychological primitives"; Lindquist et al., 2012). In this view, humans are typically in fluctuating states of pleasant or unpleasant arousal ("core affect"; Russell & Feldman Barrett, 1999; Lindquist, 2013), which can be influenced by external stimuli. Emotional arousal could thereby be a "common currency" to compare different stimuli or events (Lindquist, 2013) and represent fundamental neural processes that underlie a variety of emotions (Wilson-Mendenhall et al., 2013).

It can fluctuate quickly – on the order of minutes (Kuppens et al., 2010) or seconds (Mikutta et al., 2012) – and has been connected to ANS activity, as measured by pupil diameter (Bradley et al., 2008) or skin conductance (Bach et al., 2010). At the brain level, emotional arousal was linked to lower alpha power, particularly over parietal electrodes (Luft & Bhattacharya, 2015; Koelstra et al., 2012). The parieto-occipital alpha rhythm, typically oscillating in the frequency range of 8 to 13 Hz, is the dominant EEG rhythm in awake adults with eyes closed (Berger,

[1] Different types of arousal have been proposed and investigated (e.g., sexual, autonomic, emotional; Russell, 1980) – also in the context of altered states of consciousness (e.g., through anaesthesia or sleep). They may share psychological (e.g., increase in sensorimotor and emotional reactivity; Pfaff et al., 2012) and physiological aspects (e.g., sympathetic activation) but are not synonymous. We here explicitly refer to arousal in the context of (the subjective experience of) emotions.

1929), where it varies with vigilance (Olbrich et al., 2009). However, in tasks of visual processing (i.e., with eyes open), parieto-occipital alpha power was linked to active attentional processes (e.g., distractor suppression; Kelly et al., 2006; Klimesch, 2012) or, more generally, to functional inhibition for information gating (Jensen & Mazaheri, 2010). Physiologically, alpha oscillations were associated with large-scale synchronization of neuronal activity (Buzsáki, 2006) and metabolic deactivation (Moosmann et al., 2003).

In sum, bodily (e.g., central, autonomic) responses interact in complex ways across situations, and activity in the brain is central for emotions and their subjective component (Barrett, 2017; Seth, 2013). As (emotional) arousal is a fundamental property not only of emotions but of subjective experience in general (Adolphs et al., 2019), an investigation of its neurophysiology, reflected in neural oscillations, is essential to understanding the biology of the mind.

Studying emotional arousal and its neurophysiology in the lab

Studies that investigated emotions or emotional arousal in laboratory environments typically used static images. For example, more emotionally arousing relative to less emotionally arousing (e.g., neutral) pictures were associated with an event-related desynchronization (i.e., decrease in power) in alpha oscillations in posterior channels (Cesarei & Codispoti, 2011; Schubring & Schupp, 2019; but cf. Uusberg et al., 2013). In a study, in which emotional arousal was induced through pictures and music, blocks of higher emotional arousal were associated with decreased alpha power compared to blocks of lower emotional arousal (Luft & Bhattacharya, 2015). However, emotion-eliciting content that is repeatedly presented in trials creates an artificial experience for participants (Bridwell et al., 2018); it hardly resembles natural human behaviour and its (neuro-)physiology, which unfold over multiple continuous timescales (Huk et al., 2018). Moreover, such presentations lack a sense of emotional continuity. External events often do not appear suddenly but are embedded in an enduring sequence, in which emotions build-up and dissipate. Real-life scenarios also include anticipatory aspects where emotional components can be amplified or even suppressed, thus rendering the relationship between the corresponding neuronal events and subjective experience more complex than the one typically studied with randomized or partitioned presentations of visual or auditory stimuli.

Virtual Reality (VR) technology – particularly immersive VR, in which the user is completely surrounded by the virtual environment – affords the creation and presentation of computer-generated scenarios that are contextually rich and engaging (Diemer et al., 2015). As more naturalistic (i.e., dynamic, interactive, and less decontextualized) experiments allow to study the brain under conditions it was optimized for (Gibson, 1979; Hasson et al., 2020), and their findings may more readily generalize to real-world circumstances and provide better models of the brain (Matusz et al., 2019; Shamay-Tsoory & Mendelsohn, 2019).

In this study, we aimed to link subjective emotional arousal with alpha power in a naturalistic (i.e., dynamic and interactive) setting. Participants completed an immersive VR experience that included virtual rollercoaster rides while their EEG was recorded. They then continuously rated their emotional arousal while viewing a replay of their previous experience (McCall et al., 2015).

Methodological challenges of naturalistic experiments

To tackle the challenges of data acquired in naturalistic settings and with continuous stimuli, we make use of recent advances in signal processing and statistical modelling: Spatial filtering methods (originally developed for brain-computer interfaces, BCIs; Blankertz et al., 2008), have recently gained popularity in cognitive neuroscience (M. X. Cohen, 2018; Zuure & Cohen, 2020). Spatial filters can be used as dimensionality reduction techniques that combine data from multiple M/EEG channels to obtain a new signal (component) with a higher signal-to-noise ratio (SNR; Lotte et al., 2018; Parra et al., 2005). The filters can be optimized in different ways, for example by solving Generalized Eigenvalue Decomposition problems, to maximize the difference in the signal of interest between experimental conditions (de Cheveigné & Parra, 2014; Rivet et al., 2009) or against signals in the neighbouring frequency ranges (Nikulin et al., 2011). Spatial filtering has also been used to analyze continuous (sparse) data collected in naturalistic experiments, for example, to find inter-subjective correlations in neuroimaging data of participants watching the same movie (Biessmann et al., 2014; Dmochowski et al., 2012; Gaebler et al., 2014) or following the same lecture (Cohen et al., 2018; Poulsen et al., 2016).

For the present experiment, two spatial filtering methods were applied to link alpha power and subjective emotional arousal: Source Power Comodulation (SPoC; Dähne et al., 2014) and Common Spatial Patterns (CSP; Blankertz et al., 2008; Ramoser et al., 2000).

SPoC is a supervised regression approach, in which a target variable (here: subjective emotional arousal) guides the extraction of relevant M/EEG oscillatory components (here: alpha power). SPoC has been used to predict single-trial reaction times from alpha power in a hand motor task (Meinel et al., 2016), muscular contraction from beta power (Sabbagh et al., 2020), and difficulty levels of a video game from theta and alpha power (Naumann et al., 2016). CSP is used to decompose a multivariate signal into components that maximize the difference in variance between distinct classes (here: periods of high and low emotional arousal). CSP is being used with EEG for BCI (Blankertz et al., 2007) or to decode workload (Schultze-Krafft, 2016). Formally, SPoC is an extension of CSP for regression-like decoding problems with a continuous target variable (Dähne et al., 2014).

In addition to M/EEG-specific spatial filtering methods, non-linear machine learning methods are suited for the analysis of continuous, multidimensional recordings from naturalistic experiments. Deep neural networks transform high-dimensional data into target output variables (here: different states of emotional arousal) by finding statistical invariances and hidden representations in the input (Goodfellow et al., 2016; LeCun et al., 2015; Schmidhuber, 2015). For time-sequential data, Long Short-Term Memory (LSTM) recurrent neural networks (RNNs) are particularly suited (Greff et al., 2017; Hochreiter & Schmidhuber, 1995, 1997). Via nonlinear gating units, the LSTM determines which information flows in and out of the memory cell in order to find long- and short-term dependencies over time. LSTMs have been successfully applied for speech recognition (Graves et al., 2013), language translation (Luong et al., 2015), or scene analysis in videos (Donahue et al., 2015), but also to detect emotions in speech and facial expressions (Wöllmer et al., 2010, 2008) or workload in EEG (Bashivan et al., 2016; Hefron et al., 2017). In comparison to other deep learning methods, LSTMs are “quick learners” (due to their efficient gradient flow) and thus suitable for the continuous and sparse data recorded under naturalistic stimulation with VR.

The present study tested the hypothesis of a negative association between alpha power and (retrospectively reported) subjective emotional arousal. Combining immersive VR and EEG, this study aimed to (1) induce variance in emotional arousal in a naturalistic setting and (2) capture the temporally evolving and subjective nature of emotional arousal via continuous ratings in order to (3) assess their link to oscillations of brain activity in the alpha frequency range. This hypothesis was tested by decoding subjective emotional arousal from alpha power using the three complementary analysis techniques SPoC, CSP, and LSTM.

Results

Participants

45 healthy young participants (22 men, M±SD: 24.6±3.1, range: 20-32 years) completed the experiment. Data from 8 participants needed to be discarded due to technical problems (n = 5) or electrode malfunctioning (n = 1); one participant discontinued the experiment and another participant reported having taken psychoactive medication. The data from 37 participants entered the analysis (17 men, age: M±SD: 25.1±3.1, range: 20-31 years). After quality assurance during the EEG preprocessing, data from 26 participants in the nomov and 19 in the mov condition entered the statistical analyses that included EEG data.

Self-reports

Questionnaires

From before (M±SD: 8.68±2.82, range: 6-17) to after the experiment (M±SD: 11.82±5.24, range: 6-29), the overall simulator sickness (e.g., nausea, headache) increased significantly ($t(36) = 3.72$, $p = .0007$). As the trait questionnaires are not the focus of this study, their results will be reported elsewhere.

Emotional arousal ratings

The retrospective emotional arousal ratings for the VR experience, averaged across all subjects and timepoints, were 46.94±12.50 (M±SD, range: 16.17-66.29) in the nomov and 50.06±12.55 (M±SD, range: 18.00-69.94) in the mov condition. Qualitatively, the emotional arousal was highest for the *Andes Coaster,* lower for the *Space Coaster*, and lowest for the break (see Figure 1).

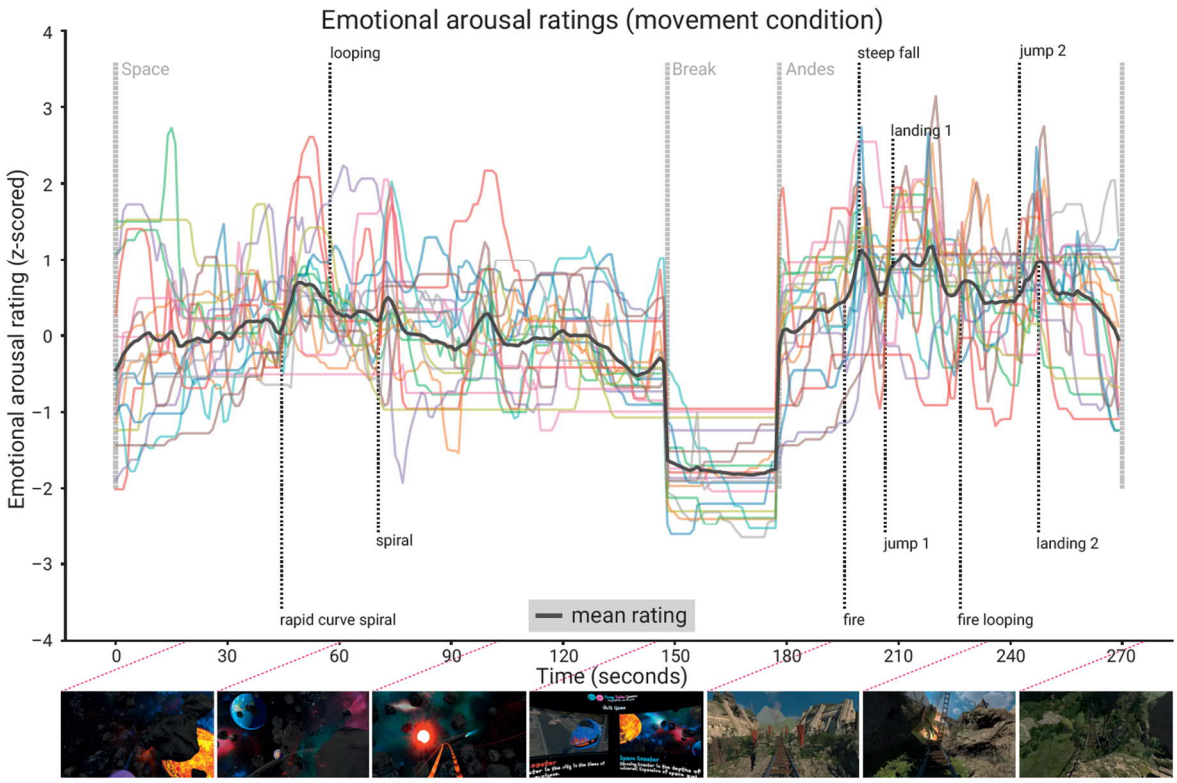

Figure 1: Emotional arousal ratings of the experience (with head movement; see Figure S1 for the ratings from the no-movement condition). Colored lines: individual participants; black line: mean across participants; vertical lines (light grey): beginning of the three phases (Space Coaster, Break, Andes Coaster); vertical lines (dark grey): manually labelled salient events (for illustration). Bottom row: Exemplary screenshots of the VR experience.

Neurophysiology

SPoC

SPoC results (Figure 2) showed that for 24/26 (92.30 %) participants in the nomov and 15/19 (78.94 %) participants in the mov condition (see Table S3 for single-participant results), the component having the highest absolute lambda value corresponded with the one that maximized the negative correlation between z and alpha power. Among the components that maximized inverse covariance, 9/26 (34.61 %) for the nomov and 6/19 (31.57 %) for the mov condition remained statistically significant ($p < .05$) after a permutation-based test (500 iterations, exact *p* values are reported in Table S3). The global mean lambda value of these components was -0.46 for the nomov (range: -1.49 - 0.08) and -0.42 for the mov condition (range: -1.61 - 0.01). The mean Spearman correlation value between the target variable z (subjective ratings) and z_{est} (estimated target variable) was significantly different from zero for both the nomov (M±SD: -.20±.10 ; range: -.36 - +.12; $t_{nomov}(25) = -9.84$; $p < .001$) and the mov condition (M±SD:-.20±.11; range: -.40 - +.007; $t_{mov}(18) = -8.16$; $p < .001$).

CSP

As assessed by a 10-fold cross-validation, the mean classification accuracies, averaged across subjects, for the model based on CSP decomposition and LDA classification were

60.83±7.40 % (M±SD; range: 47.22 - 77.78 %) for the nomov, and 60.76±6.58 % (M±SD; range: 48.33 - 71.67 %) for the mov condition. Both were significantly above chance level ($t_{nomov}(25) = 7.47$, $p_{nomov} < .001$; $t_{mov}(18) = 7.12$, $p_{mov} < .001$). At the single-subject level, the classification accuracy was significantly above chance level ($p < .05$; assessed with an exact, two sided binomial test against chance level, i.e. 50 % of samples correctly classified) for 17/26 (65.39 %) participants in the nomov, and for 12/19 (63.16 %) participants in the mov condition (see Table S3 for single-participant results). The spatial patterns yielded by the CSP decomposition are shown in Figure 4 (across participants) and in Figure S2 (individual participants). Corresponding alpha power sources (located via eLORETA) are shown In Figure 5.

LSTM

After a random search over a constrained range of hyperparameters (HPs), we extracted the best individual HP set per subject (see Supplementary Material for the list of best HPs per condition). The mean classification accuracy was 59.42±4.57 % (M±SD; range: 52.22 % - 68.33 %) for the nomov, and 61.29±4.5 % (M±SD; range: 53.89 % - 71.11 %) for the mov condition. Both were significantly above chance level ($t_{nomov}(25) = 10.82$, $p_{nomov} < .001$; $t_{mov}(16) = 10.51$, $p_{mov} < .001$). At the single-subject level, the classification accuracy was significantly above chance level for 16/26 (61.54 %) participants in the nomov condition, and for 16/19 (84.21 %) participants in the mov condition (same test as for CSP results; see Table S3 for single-participant results).

Comparison of model performances

As an illustration of the prediction behaviour across all three models in one participant (with high performance for all three decoding approaches: $rank_{SPoC} = 3$, $rank_{CSP} = 2$, $rank_{LSTM} = 1$), see Figure 2. Correlations of performances across models and experimental conditions are shown in Figure 6. The (positive) correlation between the two binary classification approaches (CSP, LSTM) was significant (after Bonferroni multiple-comparison correction), irrespective of the experimental condition (nomov, mov), meaning that subjects who could be better classified with CSP also yielded better results in the LSTM-based classification. In a repeated-measures ANOVA testing for differences in the accuracies of the two binary classification models (CSP, LSTM) and the two conditions (nomov, mov), none of the effects was significant: neither the main effect *model* ($F(1,17) = 0.02$, $p = .904$) nor the main effect *condition* ($F(1,17) = 0.72$, $p = .408$) or their interaction ($F(1,17) = 1.59$, $p = .225$). For a further comparison of the performances of the classification approaches, the respective confusion matrices are depicted in Figure 3 (average across the subjects per condition and model).

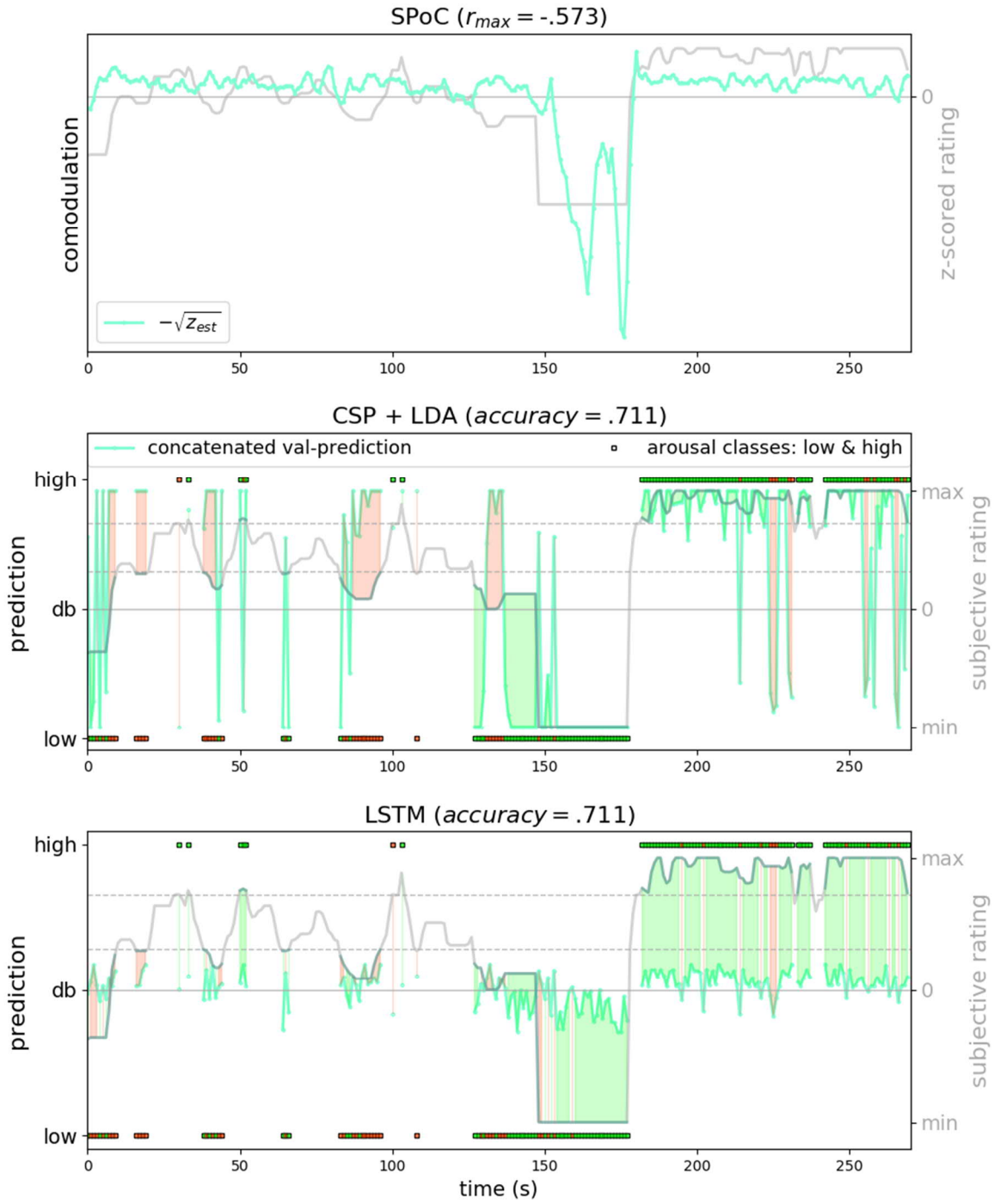

Figure 2: Predictions (turquoise line, dots) across models trained on the data of one participant in the movement condition (SPoC normalized negative z_{est}, *here comodulation;* CSP: posterior probability; LSTM: *tanh* output). *Top row*: Highest negatively correlating SPoC component (for visualization we depict the z-score of the rating and of the negative square root of z_{est}). *Middle and lower row:* Model predictions on validation sets (across the cross-validation splits) for CSP and LSTM, respectively. The gray curvy line in each panel indicates the continuous subjective rating of the participant. Horizontal dotted lines indicate the class borders. The area between these lines is the mid-tercile which was discarded for CSP and LSTM analyses. Class membership of each rating sample (1-s) is indicated by the circles at the top and bottom of the rating. A model output falling under or above the decision boundary (db) indicates the model prediction for one over the other class, respectively. The correct or incorrect prediction is indicated

by the color of the circle (green and red, respectively), and additionally colour-coded as area between model-output (turquoise) and rating.

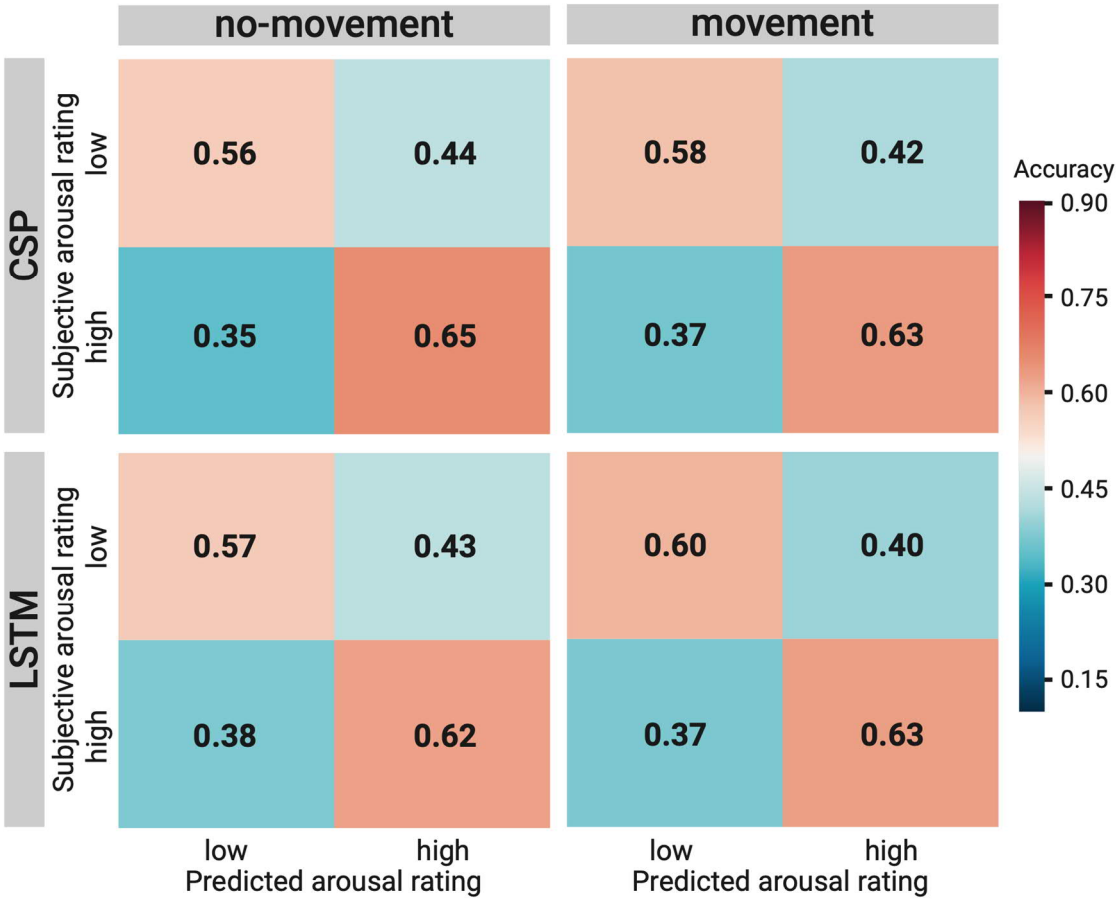

Figure 3: Confusion matrices of the classification accuracies for higher and lower self-reported emotional arousal using LSTM (*lower row*) and CSP (*upper row*) in the condition without (*left column*) and with (*right column*) head movement.

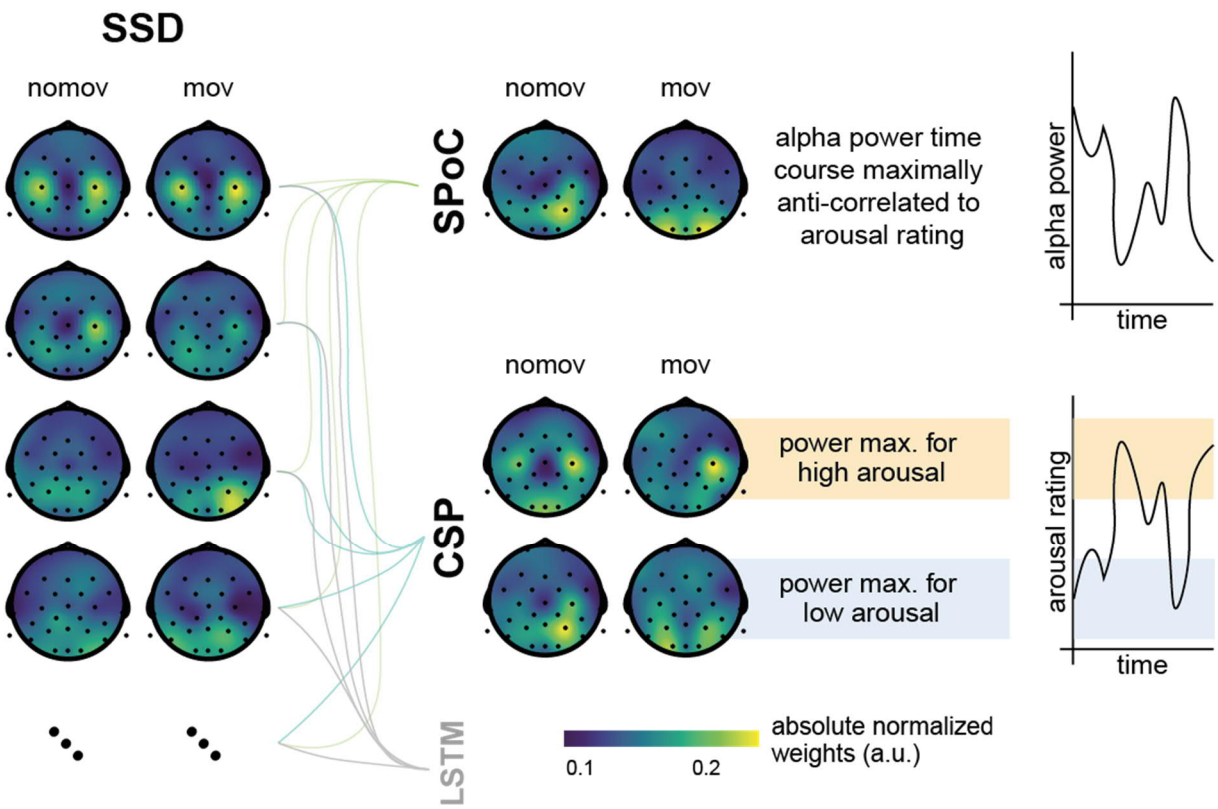

Figure 4: Spatial patterns resulting from SSD, SPoC, and CSP decomposition. Colors represent absolute normalized pattern weights (inverse filter matrices) averaged across all subjects per condition (nomov: without head movement, mov: with head movement). Before averaging, the pattern weight vectors of each individual subject were normalized by their respective L2-norm. To avoid cancellation due to the non polarity-aligned nature of the dipolar sources across subjects, the average was calculated from the absolute pattern weights. SSD allows the extraction of components with a clearly defined spectral peak in the alpha frequency band. Shown are the patterns associated with the four SSD components that yielded the best signal-to-noise ratio (*left column*). The SSD filtered signal was the input for the decoding approaches SPoC, CSP, and LSTM: SPoC adds a spatial filter, optimizing the covariance between the continuous emotional arousal ratings and alpha power. Shown here is the pattern of the component which yielded the strongest *negative* relation (according to our hypothesis of an inverse relationship between emotional arousal and alpha power). CSP decomposition yielded components with maximal alpha power for low-arousing epochs and minimal for high-arousing epochs (*bottom row in the CSP panel*) or vice versa (*upper row in the CSP panel*). The high correspondence between the patterns resulting from SPoC and the CSP patterns associated with maximum power for low arousing episodes reflects the hypothesized negative relationship between alpha power and emotional arousal. (NB: As the LSTM results cannot be topographically interpreted, they are not depicted here.)

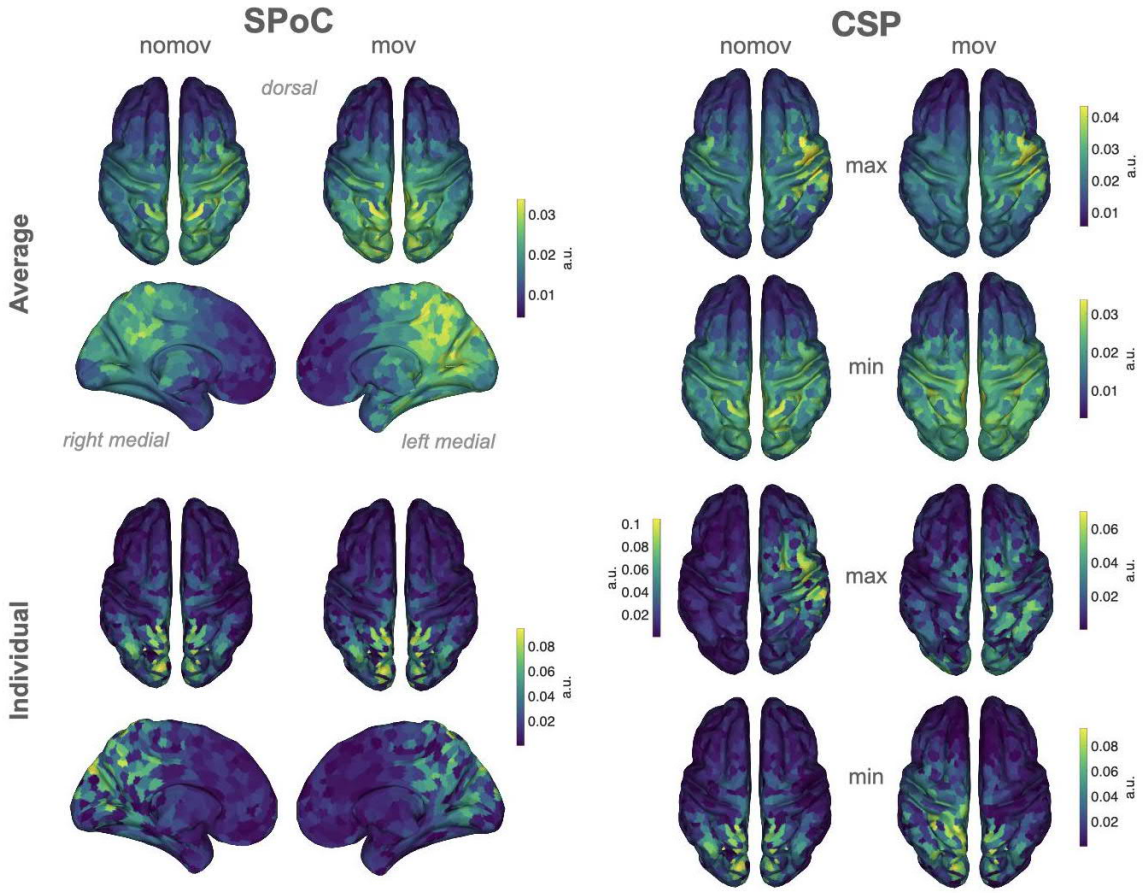

Figure 5: The projection of SPoC and CSP components in source space confirms the link between emotional arousal and alpha oscillations in parieto-occipital regions. Colors represent the inversely modelled contribution of the cortical voxels to the respective spatial pattern yielded by SPoC or CSP. We applied the same normalization and averaging procedures as for the topoplots in *Figure 4*. *Upper row*: averaged across all subjects per condition (nomov, mov). *Lower row*: patterns of one individual (the same as in Figure 2).

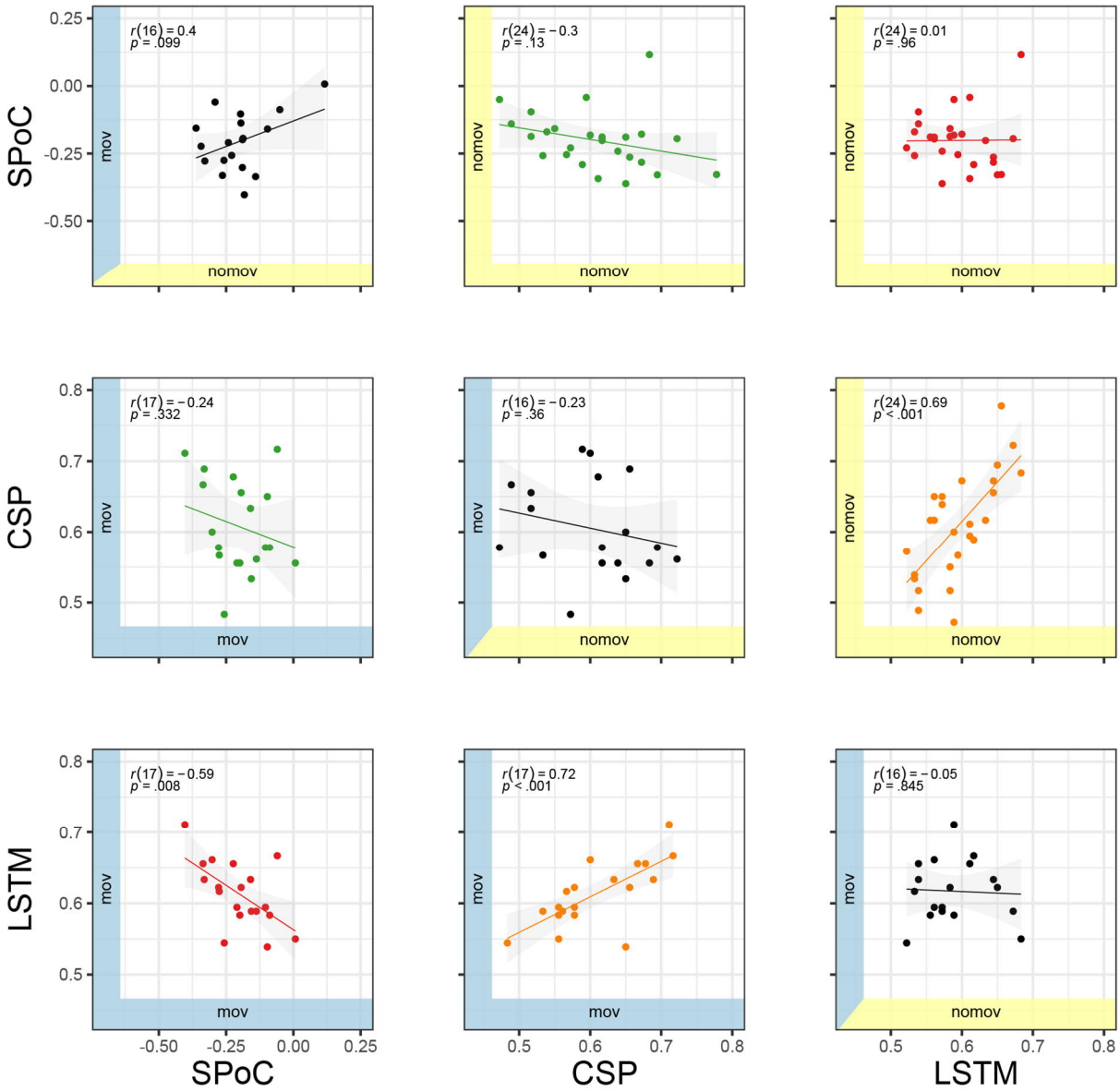

Figure 6: Correlation of performance across methods (SPoC, CSP, LSTM) and conditions (nomov: no-movement, mov: movement). The model performance metrics are classification accuracy (CSP and LSTM) and correlation coefficients (SPoC; NB: based on our hypothesis of an inverse relationship between emotional arousal and alpha power, more negative values indicate better predictive performance). Plots above and below the diagonal show data from the nomov (yellow axis shading, *upper right*) and the mov (blue axis shading, *lower left*) condition, respectively. Plots on the diagonal compare the two conditions (nomov, mov) for each method. In the top left corner of each panel, the result of a (Pearson) correlation test is shown. Lines depict a linear fit with the .95 confidence interval plotted in grey.

Discussion

The general aim of this study was to capture the dynamic relationship between subjective experience and neurophysiology under naturalistic stimulation using immersive VR. The hypothesized link between EEG alpha power and self-reported emotional arousal could be confirmed by relating alpha power components to continuous retrospective ratings of emotional arousal (using SPoC) as well as by decoding states of higher and lower emotional arousal from them (using CSP and LSTMs), particularly in parieto-occipital regions. In addition to extending our knowledge about the functional anatomy of emotional arousal, these findings support previous results from classical studies and confirm them under more naturalistic conditions. They thereby pave the way for real-world scenarios and applications.

Physiological and psychological concomitants of emotional arousal

In studies with event-related stimulation or block designs, more relative to less emotionally arousing images, videos, and sounds were associated with event-related decreases in alpha power, predominantly over parieto-occipital electrodes (Cesarei & Codispoti, 2011; Luft & Bhattacharya, 2015; Schubring & Schupp, 2019; Uusberg et al., 2013; Koelstra et al., 2012). While such stimuli provide a high degree of experimental control in terms of low-level properties and presentation timings, the emotional experience under event-related stimulation differs substantially from the emotional experience in real-life settings, which is perceptually complex, multisensory, and continuously developing over time.

Our results provide evidence that the neural mechanisms reflected in modulations of alpha power – particularly in parieto-occipital regions – also bear information about the subjective emotional state of a person undergoing an immersive and emotionally arousing experience. They thereby suggest that findings from event-related, simplified stimulation generalize to more naturalistic (i.e., dynamic and interactive) settings. Brain activity in parietal cortices was also linked to emotional processing in fMRI studies (e.g., Lettieri et al., 2019).

Paralleling the idea of emotional arousal being a dimension of “core affect” (Russell & Feldman Barrett, 1999) and a psychological primitive that underlies many mental phenomena, also alpha oscillations have been connected to various psychological “core processes”: For instance, modulations of alpha power were linked to attention (Van Diepen et al., 2019) and memory (Klimesch, 2012). More generally, neural oscillations in the alpha frequency range were suggested to serve functional inhibition of irrelevant sensory input (Jensen & Mazaheri, 2010; cf. Foster & Awh, 2019) and to code for the location and the timing of task-relevant stimuli (Foster et al., 2017). Such processes can be functionally linked to emotional arousal: During emotionally arousing experiences, preferred and enhanced processing of relevant sensory stimuli (e.g., indicating potential threats) is an adaptive behavior. In line with this, modulations of alpha oscillation over parietal sensors have been linked to threat processing (Grimshaw et al., 2014). Variations in emotional arousal and alpha power may, thus, have guided attention and memory formation also in our experiment: During particularly arousing parts of the rollercoaster, participants may have directed their attention to specific parts of the visual scene, for example, to foresee the end of the looping. Moreover, our inverse modelling (Figure 5) has also localized

arousal-related alpha sources in sensorimotor cortices, which could correspond to somatic experiences typically associated with rollercoasters.

VR as a step towards a real-world neuroscience

More naturalistic experimental stimulation, for example using immersive VR, allows to test the brain under conditions it was optimized for and thereby improve the discovery of neural features and dynamics (Gibson, 1979; Hasson et al., 2020). Findings from naturalistic studies can test the real-world relevance of results obtained in highly controlled, abstract laboratory settings (Matusz et al., 2019; Shamay-Tsoory & Mendelsohn, 2019). Challenges of using VR for more naturalistic research designs are the creation of high-quality VR content, more complex technical setups, and discomfort caused by the immersion into the virtual environment (Pan & Hamilton, 2018; Vasser & Aru, 2020). In our study, only one participant stopped the experiment because of cybersickness. This low number may result from the relatively short length of the VR experience (net length: <20 min) and the professionally produced VR stimulation. Shorter exposure times (Rebenitsch & Owen, 2016) and experiences that elicit stronger feelings of presence have been associated with lower levels of cybersickness (Weech et al., 2019).

Combining EEG with VR provides additional challenges: the SNR can decrease due to mechanical interference of the VR headset with the EEG cap and due to movement artifacts when the participant interacts with the virtual environment (e.g., head rotations). To ensure high data quality, we applied multiple measures to prevent, identify, reject, or correct artifacts in the EEG signal (see Methods section for details). Ultimately, the performance of all three decoding models did not differ significantly for both conditions (nomov, mov). We suggest that, with appropriate quality assurance during data acquisition and analysis (leading to more data rejection/correction for mov than for nomov), EEG can be combined with immersive VR and free head movements. Other studies of mobile brain imaging, even recording outdoors and with full-body movements, came to similar conclusions (Debener et al., 2012).

Evaluating EEG data from naturalistic experiments using complementary methods

Each of the applied decoding approaches allows for different insights and interpretations, but overall, they yield converging results.

SPoC and CSP

SPoC and CSP share advantages that are common to most spatial filtering methods based on Generalized Eigenvalue Decomposition, namely precise optimization policies, high speed and interpretability (see *Introduction*). The similarity between the two approaches and their interpretability becomes apparent in the resulting spatial patterns: the normalized and averaged SPoC topoplots and source localizations in both conditions (nomov, mov) resemble the ones extracted via CSP to maximize power for the low-arousal epochs of the experience (Figures 4 and 5). SPoC and CSP solve a similar problem here: extracting components whose power is minimal during states of high emotional arousal and maximal during states of low arousal.

This indicates that SPoC and CSP exploited similar spatial informational patterns in the input data. However, the datasets handed to the SPoC and CSP models were not identical. For the CSP analysis, only the upper and lower extreme of the arousal ratings were included (i.e. ⅔ of the data), while epochs with medium arousal ratings (i.e., ⅓ of the data) were excluded, whereas SPoC was trained on the full continuous datastream. There are two potential explanations for the observation that SPoC and CSP nevertheless yield similar spatial patterns: either the most relevant information was encoded in the most extreme parts of the experience, or there is a truly linear relationship between alpha power and emotional arousal that can be queried on all parts of this spectrum ranging from low to high emotional arousal.

We observed a high degree of variability across participants (see Table S1) with less than 70 % of participants for CSP (and less than 35 % for SPoC) yielding significant results on the single-subject level. This variability reflects the difficulty of some features and classifiers to generalize across subjects, which has been reported in the BCI literature (Krusienski et al., 2011; Nurse et al., 2015). LSTM, as a non-linear model class that is less constrained in the choice of its features, can complement SPoC and CSP.

LSTM

Despite having recently gained more attention with the fast progress of deep learning (e.g., more efficient hardware and software implementations), LSTMs still need to stand up to well-established models such as CSP for EEG analysis. We found that the LSTM can extract features from neural input components that reflect changes in subjective emotional arousal and that the accuracy of its predictions in both conditions (nomov, mov) matched closely the ones of CSP (see Figures 3 and 6). It is noteworthy that for the CSP model, the (LDA-based) classification rested on narrowly defined spectral features of the signal while for the LSTM model, the input was the signal in the time-domain and the feature selection process was part of the model fitting. The strong correlation between the predictions of the two models suggests that the LSTM extracts similar information as the CSP to make its prediction, namely power. Higher accuracies may be achievable with LSTM models due to their non-convex optimization landscape. However, in our two-step hyperparameter search, we found that for each subject a range of different hyperparameter settings led to similar prediction accuracies (see Supplementary Material). Model ensembles, although computationally demanding, could further increase the robustness of the estimates (Opitz & Maclin, 1999; Rokach, 2010; Dietterich, 2000). Although it is often stated that deep learning models require large datasets (for an empirical perspective, see Hestness et al., 2017), our model, with its architecture of 1-2 LSTM layers followed by 1-2 fully connected layers, converged in less than 200 training iterations on a relatively small dataset. This quick convergence is partly due to the fast gradient-flow through the memory cell of the LSTM during the weight update, which is an additional advantage of the LSTM over other RNNs (Doetsch et al., 2014; Hochreiter & Schmidhuber, 1997). Additionally, the spatial-spectral filtering in our study (i.e., SSD-based extraction of narrow-band alpha components) may have eased the training of the LSTM. With more data, an LSTM could be trained on raw data or longer segments of the EEG to preserve more of the continuous structure and ultimately exploit its central property of detecting long-term dependencies in the input.

In contrast to SPoC and CSP, we did not compute explanatory topoplots or sources from the LSTM, since the analysis of predictions on input level in non-linear deep learning models constitutes a challenge in itself (i.e., “black box” problem of deep learning). However, “*explainable*

artificial intelligence" (XAI) is an active area of research in machine learning, aiming to open this "black box". For EEG, there are attempts to create topologically informative maps in the signal space that explain the decision of simple shallow neural networks (Sturm et al., 2016). Also for the more complex LSTM model, XAI methods were applied, for example, on text data (Arras et al., 2017; see also Lapuschkin, 2019). However, exploring and validating these approaches on our data was beyond the scope of this study.

In summary, we find that SPoC, CSP, and LSTM can be used to decode subjective emotional arousal from EEG acquired during a naturalistic immersive VR experience. The source of the alpha oscillations could be localized in parieto-occipital regions.

Compared to other EEG decoding paradigms (e.g., lateralized motor imagery; Herman et al., 2008), the accuracy of our models was relatively low. This may be a consequence of (1) the fast changing events in the VR experience (particularly the rollercoasters), (2) the asynchronicity of the two data streams as participants retrieved their emotional states from memory in retrospective ratings, (3) the generally high inter-individual variability in the interpretability of subjective self-reports (Blascovich, 1990), and (4) the "single-trial" study design and its relatively short time series. With respect to (1)-(3), people's memory for feelings and events is susceptible to distortions and biases (Kaplan et al., 2016; Levine & Safer, 2002). Following McCall et al. (2015), we elicited the memory recall by showing participants an audiovisual replay of their experience from their own perspective in the VR headset while recording continuous ratings. This aimed to minimize biases related to the point of view (Berntsen & Rubin, 2006; Marcotti & Jacques, 2018) or time scale (e.g. Fredrickson & Kahneman, 1993) during recall (as discussed in McCall et al., 2015). Lastly, while our research aimed to explore the role of the alpha frequency band in the appraisal of emotional arousal (see Introduction), higher frequencies could carry additional information about the phenomenon leading to better model predictions. However, higher frequency bands also include non-neural (e.g., muscle activity-related) signals, limiting the interpretability of those results.

Limitations

Our study has limitations that need to be considered when interpreting the results:

While being engaging, emotionally arousing and tolerable for the subjects, the commercial content used for stimulation did not provide access to the source code in order to control and extract stimulus features (e.g., height or speed of the rollercoasters). In general, creating high-quality VR content is a challenge for research labs, but there are recent efforts to provide toolboxes that facilitate customized VR development (e.g., Underwood Project; Schofield & McCall, 2020) and scientific experimentation in VR (e.g., Grübel et al., 2017; Brooks et al, 2019).

The length of the experience was chosen to minimize habituation to the stimulus and inconvenience caused by the recording setup (EEG electrodes and VR headset). This led to relatively short recording times per subject and condition. Data sparsity, however, is challenging for decoding models, which need a sufficient amount of data points for model training and evaluation, where especially larger training sets lead to more robust predictions (Hestness et al., 2017). We used cross-validation, which is commonly applied in scenarios of limited data, to achieve a trade-off between training and validation data (Varoquaux et al., 2017). Nevertheless, the models and results can be expected to perform more robustly with more training data.

Finally, emotional arousal is a multi-faceted mind-brain-body phenomenon that involves the situated organism and its interaction with the environment. The training data for multivariate models such as the LSTM can include other modalities, such as peripheral physiological (e.g., HR, GSR) or environmental (e.g., optical flow) features. Naturalism can be further increased by sensorimotor interaction (beyond head movements) in immersive VR (McCall et al., 2015) or by mobile EEG studies in real-world environments (Debener et al., 2012), which, however, poses further challenges to EEG signal quality (Gwin et al., 2010).

Conclusion

We conclude that different levels of subjectively experienced emotional arousal can be decoded from neural information in naturalistic research designs. We hope that combining immersive VR and neuroimaging not only augments neuroscientific experiments but also increases the generalizability and real-world relevance of neuroscientific findings.

Methods and Materials

Participants

45 healthy young participants were recruited via the participant database at the Berlin School of Mind and Brain (an adaption of ORSEE; Greiner, 2015). Previous studies on the relationship between emotional arousal and neural oscillations reported samples of 19-32 subjects (e.g., Koelstra et al., 2011; Luft & Bhattacharya, 2015). We recruited more participants to compensate for anticipated dropouts due to the VR setup and to ensure a robust estimate of the model performances. Inclusion criteria were right-handedness, normal or corrected-to-normal vision, proficiency in German, no (self-reported) psychiatric or neurological diagnoses in the past ten years, and less than 3 hours of experience with VR. Participants were requested to not drink coffee or other stimulants one hour before coming to the lab. The experiment took ~2.5 hours, and participants were reimbursed with 9 € per hour. They signed informed consent before their participation, and the study was approved by the Ethics Committee of the Department of Psychology at the Humboldt-Universität zu Berlin.

Setup, stimuli, and measures

The experiment was conducted in a quiet room, in which the temperature was kept constant at 24°C.

Neurophysiology / EEG

30 channels of EEG activity were recorded in accordance with the international 10/20-system (Sharbrough et al., 1991) using a mobile amplifier (LiveAmp32) and active electrodes (actiCap; both by BrainProducts, Gilching, Germany). Two additional electrooculogram (EOG) electrodes were placed below and next to the right eye to track eye movements. Data were sampled at 500 Hz with a hardware-based low-pass filter at 131 Hz and referenced to electrode FCz. The amplifier was placed on a high table in the back of the participant to minimize the pull on electrode cables and provide maximal freedom for head movements. The VR headsetwas placed carefully on top of the EEG cap, and impedances were brought below 10 kΩ. With the same amplifier, electrocardiography and galvanic skin responses were additionally acquired. These peripheral physiological data and the inter-individual differences in interoceptive accuracy are beyond the scope of this paper, and their results will be reported elsewhere.

Virtual Reality (VR) head-mounted display (HMD)

An HTC Vive HMD (HTC, New Taipei, Taiwan) and headphones (AIAIAI Tracks, ApS, Copenhagen, Denmark) were placed on top of the EEG cap using small, custom-made cushions to avoid pressure artifacts and increase comfort. The HTC Vive provides stereoscopy with two 1080 x 1200-pixel OLED displays, a 110° field-of-view, and a frame rate of 90 Hz. The user's head position is tracked using infrared light, accelerometry, and gyroscopy. Head movements were recorded by adapting scripts from https://github.com/Omnifinity/OpenVR-Tracking-Example/.

Immersive VR experience / stimulation

Stimulation comprised two commercially available rollercoaster rides ("Russian VR Coasters" by Funny Twins Games, Ekaterinburg, Russia, on Steam) that were separated by a 30-s break (during which participants kept their eyes open and looked straight): the "Space" rollercoaster, a 153-s ride through planets, asteroids, and spaceships and the "Andes" rollercoaster, a 97-s ride through a mountain scenery (for more details, see Supplementary Material). The two rollercoaster rides were commercially available on Steam. The rollercoasters were selected for their length (i.e., not leading to physical discomfort by uninterruptedly wearing the HMD for too long) and content (i.e., inducing variance in emotional arousal). The experience (i.e., the sequence "Space"-break-"Andes") was kept constant across participants.

Self-reports

Questionnaires

At the beginning of the experiment, participants completed two arousal-related questionnaires: (1) the "Trait" subscale of the "State-Trait Anxiety Inventory" (STAI-T; Spielberger, 1983, 1989) and (2) the "Sensation Seeking" subscale of the "UPPS Impulsive Behaviour Scale" (UPPS; Schmidt et al., 2008; Whiteside & Lynam, 2001). Before and after the experiment, participants completed a customized version of the "Simulator Sickness Questionnaire" (SSQ, Bouchard et al., 2017) comprising three items from the nausea (general discomfort, nausea, dizziness) and three items from the oculomotor subscale (headache, blurred vision, difficulty concentrating) to capture potential VR side effects (Sharples et al., 2008). After the experiment, participants also rated the presence and valence of their experience (the results will be reported elsewhere).

Emotional arousal

After each VR experience, participants watched a 2D recording (recorded using OBS Studio, https://obsproject.com/) of their experience on a virtual screen (SteamVR's "view desktop" feature), that is, without removing the HMD. They recalled and continuously rated their emotional arousal by turning a dial (PowerMate USB, Griffin Technology, Corona, CA, USA; sampling frequency: 50 Hz), with which they manipulated a vertical rating bar, displayed next to the video, ranging from low (0) to high (100) in 50 discrete steps (McCall et al., 2015; see Figure 7B). The exact formulation was "When we show you the video, please state continuously how emotionally arousing or exciting the particular moment during the VR experience was" (German: "Wenn wir dir das Video zeigen, gebe bitte durchgehend an, wie emotional erregend, bzw. aufregend der jeweilige Moment während der VR Erfahrung war"). To present the playback video and the rating bar a custom script written in Processing (v3.0) was used.

Procedure

Participants came to the lab and filled in the pre-test questionnaires. After the torso and limb electrodes had been attached, participants completed a heartbeat guessing task (Schandry, 1981) to assess inter-individual differences in interoceptive accuracy (the results of peripheral physiology and interoception will be reported elsewhere). Then, the EEG cap was attached, and the HMD was carefully placed on top of it. To prevent or minimize (e.g., movement-related)

artefacts, customized cushions were placed below the straps of the VR headset to reduce the contact with the EEG sensors. In addition, the VR experience took place while seated and without full body movements (participants were asked to keep their feet and arms still during the recordings). A white grid was presented in the HMD to ensure that the participants' vision was clear. They then completed a 10-min resting-state phase (5 min eyes open, 5 min eyes closed), before experiencing the first VR episode, which consisted of the two virtual rollercoaster rides and the intermediate break: First the "Space" and then, after the break, the "Andes" rollercoaster. In the subsequent rating phase, they recalled and continuously rated their emotional arousal while viewing a 2D recording of their experience. Importantly, each participant completed the VR episode (plus rating) twice: once while not moving the head (*nomov*) and once while freely moving the head (*mov*) during the VR experience. The sequence of the movement conditions was counterbalanced across participants (n = 19 with nomov condition first). At the end of the experiment, participants completed two additional questionnaires (the SUS and the questionnaire on subjective feelings of presence and valence during the virtual rollercoaster rides) before they were debriefed.

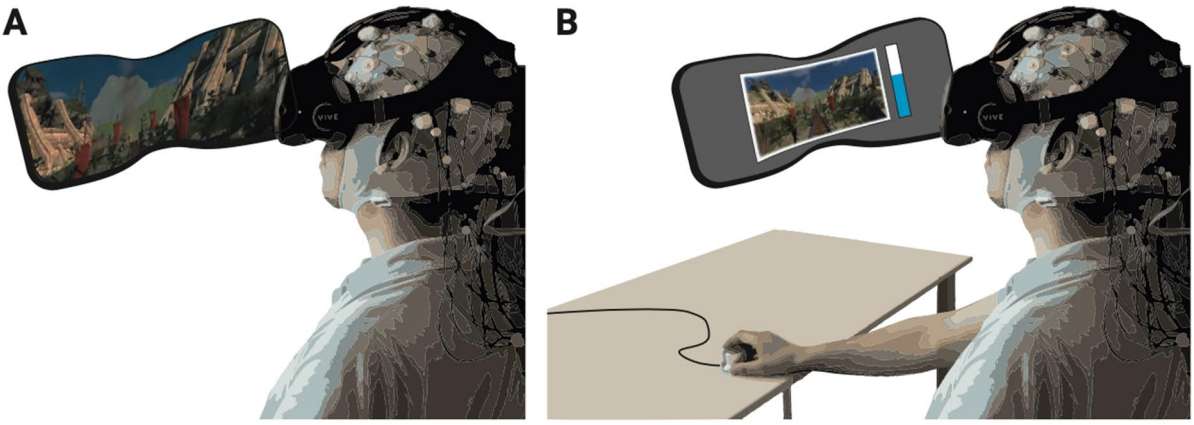

Figure 7: Schematic of experimental setup. (A) The participants underwent the experience (two rollercoasters separated by a break) in immersive VR, while EEG was recorded. (B) They then continuously rated the level of emotional arousal with a dial viewing a replay of their experience. The procedure was completed twice, without and with head movements.

Data analysis

To exclude effects related to the on- or offset of the rollercoasters, data recorded during the first and the last 2.5 s of each rollercoaster were removed and the inter-individually slightly variable break was cropped to 30 s. The immersive VR experience that was analysed thus consisted of two time series of 270 s length per participant (nomov and mov).

Self-reports

Questionnaires

Inter-individual differences as assessed by the trait questionnaires were not the focus of this study, and their results (together with the peripheral physiological and interoception data) will

be reported elsewhere. The sum of the simulator sickness ratings before and after the experiment was compared using a paired *t*-test.

Emotional arousal

Emotional arousal ratings were resampled to 1 Hz by averaging non-overlapping sliding windows, yielding one arousal value per second. For the classification analyses, ratings were divided by a tertile split into three distinct classes of arousal ratings (low, medium, high) per participant. For the binary classification (high vs low arousal), the medium arousal ratings were discarded.

Neurophysiology

Preprocessing

EEG data were preprocessed and analyzed with custom MATLAB scripts building on the EEGLAB toolbox (v13.5.4b; Delorme & Makeig, 2004). The preprocessing steps were applied separately for data recorded during the nomov and mov conditions (i.e., without and with head movement). Continuous data were downsampled to 250 Hz (via the 'pop_resample.m' method in EEGLAB) and PREP pipeline (v0.55.2; Bigdely-Shamlo, Mullen, Kothe, Su, & Robbins, 2015) procedures were applied for detrending (1 Hz high-pass filter, Hamming windowed zero-phase sinc FIR filter, cutoff frequency (-6 dB): 0.5 Hz, filter order: 827, transition band width: 1 Hz), line-noise removal (line frequency: 50 Hz), robust referencing to average, and detection as well as spherical interpolation of noisy channels. Due to the relatively short lengths of the time series, the default fraction of bad correlation windows (parameter 'badTimeTheshold', used to demark bad channels) was increased to 0.05. For all other parameters, PREP's default values were kept. On average, 2.08 and 2.47 channels per subject were interpolated in the nomov and mov condition, respectively. Data remained high-pass filtered for the further steps of the analysis. Retrospective arousal ratings were added to the data sets, labelling each second of data with an associated arousal rating used as target for the later classification and regression approaches.

ICA decomposition was used to identify and remove EEG artifacts caused by eye movements, blinks, and muscular activity. To facilitate the decomposition, ICA projection matrices were calculated on a subset of the data from which the noisiest parts had been removed. To this end, a copy of the continuous data was split into 270 epochs of 1 s length. Epochs containing absolute voltage values >100 µV in at least one channel (excluding channels that reflected eye movements, i.e., EOG channels, Fp1, Fp2, F7, F8) were deleted. Extended infomax (Lee et al., 1999) ICA decomposition was calculated on the remaining parts of the data (after correcting for rank deficiency with a principal component analysis). Subjects with >90 to-be-deleted epochs (33% of the data) were discarded from further analyses (nomov: n = 5; mov: n = 10). Artifactual ICA components were semi-automatically selected using the SASICA extension (Chaumon et al., 2015) of EEGLAB and visual inspection. On average, 13.41 (nomov) and 10.31 (mov) components per subject were discarded. The remaining ICA weights were back-projected onto the continuous time series.

Dimensionality reduction: SSD in the (individual) alpha frequency range

Our main hypothesis was that EEG-derived power in the alpha frequency range allows the discrimination between different states of arousal. To calculate alpha power, we adopted spatio-

spectral decomposition (SSD; Nikulin et al., 2011) which extracts oscillatory sources from a set of mixed signals. Based on Generalized Eigenvalue Decomposition, it finds the linear filters that maximize the signal in a specific frequency band and minimize noise in neighbouring frequency bands. A pre-processing with SSD has been previously shown to increase classification accuracy in BCI applications (Haufe, Dähne, et al., 2014). The alpha frequency range is typically fixed between 8 and 13 Hz. The individual alpha peak frequency, however, varies intra- and inter-individually, for example, with age or cognitive demand (Haegens et al., 2014; Mierau et al., 2017). To detect each participant's individual peak of alpha oscillations for the SSD, (1) the power spectral density (PSD) of each channel was calculated using Welch's method ($segment\ length\ =\ 5s\ *\ sampling\ frequency$ [i.e., 250 Hz] with 50% overlap) in MATLAB (*pwelch* function). (2) To disentangle the power contribution of the 1/f aperiodic signal from the periodic component of interest (i.e., alpha), the MATLAB wrapper of the FOOOF toolbox (v0.1.1; Haller et al., 2018; frequency range:]0-40] Hz, peak width range: 1-12 Hz, no minimum peak amplitude, threshold of 2 SD above the noise of the flattened spectrum) was used. The maximum power value in the 8-13 Hz range was considered the individual alpha peak frequency α_i, on which the SSD bands of interest were defined (bandpass signal $\alpha_i \pm 2$ Hz, bandstop noise $\alpha_i \pm 3$ Hz, bandpass noise $\alpha_i \pm 4$ Hz).

The entire procedure was separately applied to the nomov and the mov condition to account for potential peak variability (Haegens et al., 2014; Mierau et al., 2017). SSD was then computed based on these peaks. A summary of the resulting individual alpha peak frequencies can be found in Table S1.

SSD Components Selection

The SSD components with sufficient alpha information (i.e., power in the alpha frequency range that exceeds the noise level) were selected with the following steps (see Figure 8):

(1) The power spectral density of a component was calculated using Welch's method ($segment\ length\ =\ 5s\ *\ sampling\ frequency$ [i.e., 250 Hz] with 50% overlap) implemented in SciPy (v. 1.4.1., Jones, Oliphant, & Peterson, 2001).

(2) The 1/f curve was fitted in the signal range between]0-40] Hz, excluding a ±4 Hz window around the individual *alpha peak frequency* α_i of the subject *i*. The 1/f curve was defined (in log scale) as $f^{-1} = log(\frac{1}{a \cdot x^b})$, where *x* is the given component in the frequency domain, a serves as stretch parameter, and b represents the slope of the 1/f curve.

(3) After fitting these parameters, the signal was detrended with respect to the estimated 1/f curve.

(4) Those components were selected, whose alpha peak power in the detrended alpha window (as defined in (1)) was (A) greater than zero plus a decision threshold, which was set to .35 $\frac{\mu V^2}{Hz}$, and (B) higher than the mean amplitude of the adjacent frequency flanks of 2-Hz width on both sides of the window, i.e. $\max power(alpha\ peak) - \max(mean\ power(flank)) \geq\ 1.45\ SD$ (after z-scoring the detrended signal). The two criteria guaranteed the selection of components with a clearly defined alpha-amplitude peak over the noise-threshold defined by f^{-1}(see Figure 8).

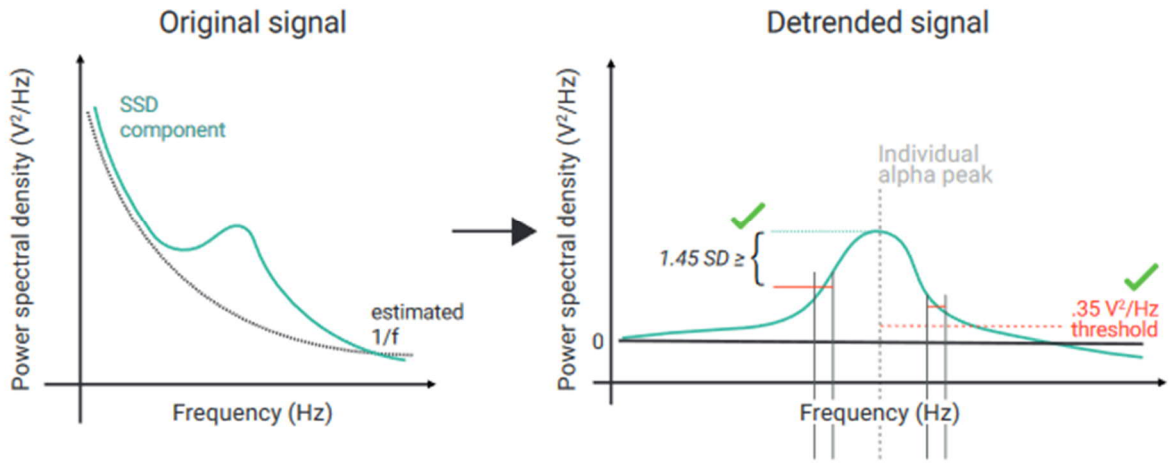

Figure 8: Schematic of the selection of individual alpha components using spatio-spectral decomposition (SSD). *(Left)* 1/f-estimation (dotted grey line) to detrend SSD components (solid turquoise line). *(Right)* After detrending the signal, components were selected, whose peak in the detrended alpha window (centred at the individual alpha peak, vertical dotted grey line) was (A) >0.35 V^2/Hz (indicated by horizontal dotted red line) and (B) higher than the bigger of the two mean amplitudes of the adjacent frequency flanks (2-Hz width).

Particularly the combination of SSD with narrow-band filtering in the alpha-frequency range lowers the probability of signal contamination elicited by artifact-related oscillations, which are typically strongest in frequency ranges above (e.g., muscular activity; Muthukumaraswamy, 2013) or below the alpha band (e.g., skin potentials, Kappenman & Luck, 2010, or eye blinks, Manoilov, 2007; for a comprehensive overview, see also Chaumon et al., 2015). Decoding models (CSP, LSTM, SPoC; described below) were trained on those subjects with at least 4 selected SSD components (26 in the nomov and 19 in the mov). On average, 7.63 of 18.81 (40.53 %) in the nomov and 5.63 of 15.22 (36.98 %) SSD components were selected in the mov condition. Importantly, SSD components had spatial topographies corresponding to occipito-parietal and fronto-central source locations, thus covering brain areas previously implicated in emotional arousal and its regulation.

Source-Power Comodulation (SPoC)

To test the hypothesis that alpha power in the EEG covaries with the level of subjective emotional arousal, SPoC[2] (v1.1.0; Dähne et al., 2014) was applied to EEG data composed of the selected SSD components and filtered around the central individual alpha peak. The supervised decomposition procedure takes the variable *z* as target, which comprised the continuous arousal ratings (z-scored; 270 s per participant). To reach the same temporal resolution as *z* (i.e., 1 Hz), EEG data were epoched into 270 consecutive segments of 1 s length. The power of a SPoC component ($\hat{s} = W^T X$, where W^T corresponds the transpose of the unmixing matrix W and X to the data matrix in SSD space) in a specific epoch (e) can be approximated by the variance of its signal within that time interval ($Var[\hat{s}](e)$Dähne et al., 2014). SPoC was separately applied to each participant, producing a number of components equal to the number of

[2] Throughout the paper, "SPoC" refers to $SPoC_\lambda$ (for details, see Dähne et al., 2014)

previously selected SSD components. The stability and significance of the extracted components was tested with a permutation approach (500 iterations): *z* values were randomly shuffled to create a surrogate target variable with randomized phase but same auto-correlation. Lambda values (*λ*, i.e., component-wise covariance between *z* and alpha power) were then calculated for each iteration to generate a naive probability density function (i.e., null-hypothesis distribution) and to estimate the probability that the lambda value that was calculated with the original target variable *z* was obtained by chance. Given our main hypothesis of an inverse relationship between alpha power and self-reported emotional arousal, we only retained, for each participant, the component with the highest negative lambda value *λ* (disregarding the p-value to avoid circularity; Kriegeskorte et al., 2009). In accordance with the primary objective of SPoC to reconstruct the target variable *z*, group-level statistics were then computed on the correlation coefficient *r* between *z* and z_{est}, which is obtained via the following formula: $z_{est} = Var[W^{(i)T}X](e)$ (where i is the index of the aforementioned most negative selected component).

Since the extracted linear spatial filters W cannot be directly interpreted (Haufe, Meinecke, et al., 2014), topographical scalp projection of the components are represented by the columns of the spatial patterns matrix A obtained by inverting the full matrix W (Figure 4).

Common Spatial Patterns (CSP)

To further test the hypothesis of a link between alpha power and subjective emotional arousal, we aimed to distinguish between the most and the least arousing phases of the experience by using features of the alpha band-power of the concurrently acquired EEG signal. We followed an approach which has successfully been used in BCI research to discriminate between event- or state-related changes in the bandpower of specific frequency ranges in the EEG signal: The Common Spatial Patterns (CSP) algorithm specifies, by means of a Generalized Eigenvalue Decomposition, a set of spatial filters to project the EEG data onto components whose bandpower maximally relates to the prevalence of one of two dichotomous states (Blankertz et al., 2008; Ramoser et al., 2000). In our case, we were interested in distinguishing moments that had been rated to be most (top tertile) and least arousing (bottom tertile).

Using the EEGLAB extension BCILAB (v1.4-devel; Kothe & Makeig, 2013), data of the selected SSD components, bandpass filtered around the individual alpha peak ± 2 Hz, were epoched in 1-s segments. This sample length was chosen to enable the extraction of neural features and corresponding changes in the subjective experience, while maximising the number of samples from the sparse datasets. Epochs with mid-level arousal ratings (middle tertile) were discarded, yielding 180 epochs (90 per class) for each subject (per movement condition). To assess the classification performance, a randomized 10-fold cross-validation procedure, a common solution for sparse training data (Bishop, 2006), was used. Per fold, a CSP-based feature model was calculated on the training data by decomposing the signal of the selected SSD components according to the CSP algorithm. A feature vector comprising the logarithmized variance of the four most discriminative CSP components (using two columns from each side of the eigenvalue decomposition matrix as spatial filters) was extracted per epoch. Data from the training splits were used to train a linear discriminant analysis (LDA) on these feature vectors (Fisher, 1936). Covariance matrices used for calculating the LDA were regularized by applying the analytic solution implemented in BCILAB (Ledoit & Wolf, 2004). The LDA model was then used to classify the feature vectors extracted from the epochs in the test split to predict the according arousal label. Average classification accuracy (defined as $1 -$

$misclassification\ rate$) over the ten folds was taken as outcome measure to assess the predictive quality of the approach. To allow a spatial interpretation of the projections, like with the SPoC components, the spatial patterns of the two most extreme CSP components (associated with the largest and smallest eigenvalue) that were used to calculate the feature vectors for the linear classification were plotted in Figure 4 (normalized and averaged across subjects per condition) and Figure S2 (per single subject and condition). Source localized patterns are shown in Figure 5.

Source Localization

Exact Low Resolution Tomography Analysis (eLORETA; Pascual-Marqui, 2007) was used to localize the sources corresponding to the component extracted via SPoC and CSP. Our pipeline was based on the work of Idaji et al. (2020), who customized the eLORETA implementation of the MEG/EEG Toolbox of Hamburg (https://www.nitrc.org/projects/meth/).

Our forward model was constructed via the New York Head model (Haufe, Meinecke, et al., 2014; Haufe & Ewald, 2019; Huang et al., 2016) with approximately 2000 voxels and by using 28 out 30 scalp electrodes (TP9 and TP10 were removed because they are not contained in the model). Crucially, we focused on dipoles perpendicular to the cortex. eLORETA was then used to construct a spatial filter for each voxel from the leadfield matrix, and respective sources were computed by multiplying the resultant demixing matrix with the spatial patterns (A) of the selected SPoC and CSP components. Inverse modelling was computed separately per participant and condition before it was averaged for each condition across all subjects (Figure 5).

Long Short-Term Memory (LSTM) Recurrent Neural Network

Deep learning models have become a useful tool to decode neural information (e.g., Agrawal et al., 2014; Khaligh-Razavi & Kriegeskorte, 2014). Applying a deep learning approach to the time series of EEG recordings (e.g., Bashivan et al., 2016) can be achieved using Long Short-Term Memory (LSTM) recurrent neural networks (Hochreiter & Schmidhuber, 1995, 1997). With their property to store and control relevant information over time, they can find adjacent as well as distant patterns in (time) sequential data. The LSTM analysis was implemented in the Python-based deep learning library *TensorFlow* (v.1.14.0; Google Inc., USA; Abadi et al., 2015; Zaremba et al., 2015).

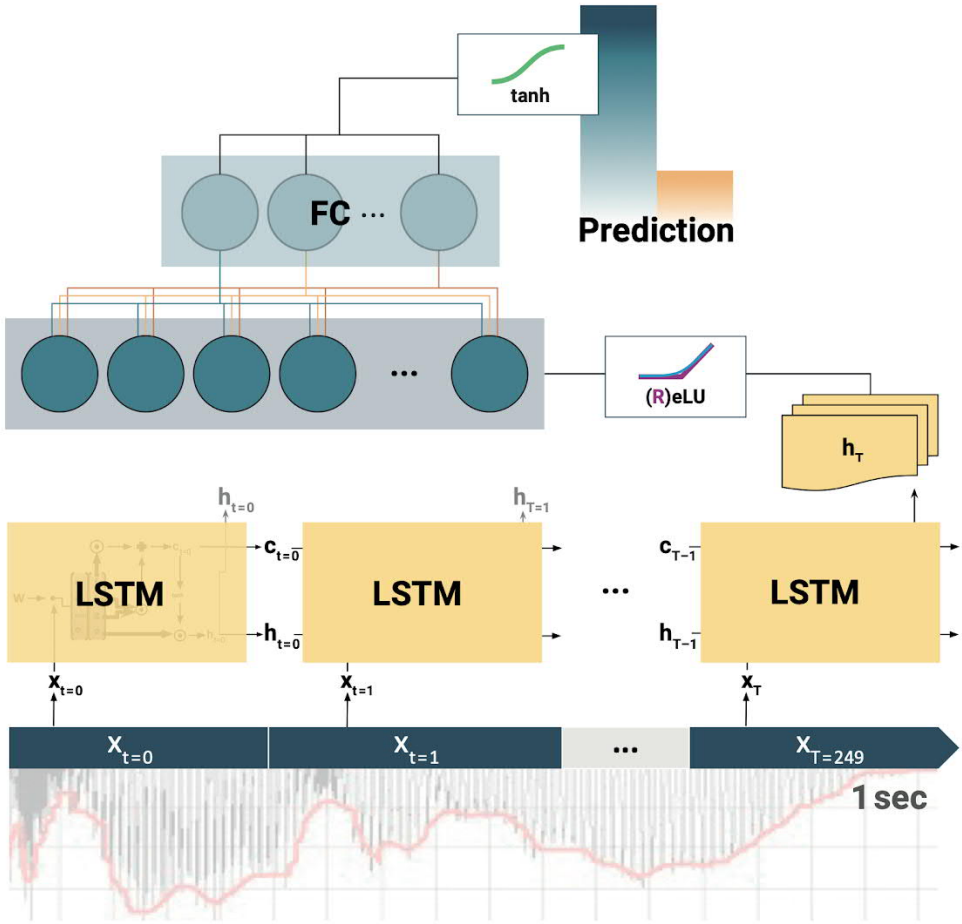

Figure 9: Schematic of the Long Short-Term Memory (LSTM) recurrent neural network (RNN). At each training step, the LSTM cells successively slide over 250 data-arrays of neural components ($x_{t=0}$, $x_{t=1}$, ... , $x_{T=249}$) corresponding to 1-s of the EEG recording. At each step t, the LSTM cell computes its hidden state h_t. Only the final LSTM output (h_T) at time-step T=249 is then fed into the following fully connected (FC) layer. The outputs of all (LSTMs, FCs) but the final layer are normalised by Rectified linear units (ReLU) or exponential linear units (ELU). Finally, the model prediction is extracted from the last FC layer via a tangens hyperbolicus (tanh). Note: depending on model architecture, there were 1-2 LSTM layers, and 1-2 FC layers.

Model Architecture and Hyperparameter Search

Deep learning models usually have a high variance due to random weight initialization, architectural choices, and hyperparameter HPs (Geman et al., 1992; but see Neal et al., 2019). We here used a two-step random search (Bergstra & Bengio, 2012) strategy in order to find optimal HPs, to reduce the model variance and make the search computationally feasible. First, a broad random search was applied on a random subset of ten subjects (20 random combinations) in each condition. Then, the two best HPs per subject were taken and applied to the datasets of all subjects. Due to time constraints and computational complexity, the HP search was limited to a predefined range of settings and the model architecture was constrained to maximal two LSTM layers followed by maximal two fully connected layers (*FC*; Hefron et al., 2017; see Figure 9). Each layer size ls_l varied between 10 and 100 nodes ($ls_l \in$ 10, 15, 20, 25, 30, 40, 50, 65, 80, 100), and a successive layer needed to be equal or smaller in size (bottleneck architecture). The output of each layer was squashed through either rectified linear units or exponential linear units, which both allow for steeper learning curves in contrast to conventional activation functions such as sigmoid nonlinearity (Clevert et al., 2016). The output of the last network layer (FC_L) was fed into a tangens hyperbolicus (*tanh*) to match the binned ratings, which were labelled with -1 or 1, respectively. We applied a mean-squared error loss to calculate the difference between the model output (i.e. the prediction) and the labels, leading to a stronger weighting of losses at the upper- or lower-class border, respectively. To control and tax too large model weights, different regularization methods (*L1*, *L2*) with different regularization strengths ($\lambda \in$ 0.00, 0.18, 0.36, 0.72, 1.44) were tested. Weights were optimized using *Adam* (learning rate: $lr \in 1e^{-2}, 1e^{-3}, 5e^{-4}$) due to its fast convergence (Kingma & Ba, 2015; see also Ruder, 2017). The number of input components (SSD, N_{comp}: N ∈ [1, 10]) was treated as HP. The specific N_{comp} neural components were successively drawn according to the order of the SSD selection.

Training procedure

The final dataset per subject was a three-dimensional tensor of size 270×250×*10 (epochs x samples x components)*. If less than 10 components were extracted for a given subject, the tensor was filled with zero-vectors. After some test runs and visual observation of the convergence behaviour of the learning progress, training iterations were set to 20 (i.e., the model ran 20 times through the whole training dataset). The 1-sec samples were fed into the LSTM in random mini-batches of size 9 (*bs* = 9), since training on batches allows for faster and more robust feature learning (Ruder, 2017), leading to the following input tensor at training step *ts*: $x_{train,ts}^{bsx250x10}$.

Statistical Evaluation

To test whether the results of the binary modelling approaches (CSP, LSTM) were statistically significant, exact binomial tests were conducted per subject and experimental condition (nomov, mov) over all 180 epochs of the respective time series (nomov, mov). To do so, for each of the binary modelling approaches (CSP, LSTM), the predictions for the single epochs in the ten test splits of the CV were concatenated to a single vector. The proportion of correct and false predictions was then compared to a null model with prediction accuracy 0.5 (chance level). To test the average (across subjects) classification accuracies of the binary models, we calculated one-sample *t*-tests, comparing the mean accuracy of the respective model for both

experimental conditions against the theoretical prediction accuracy of a random classifier (0.5). To test whether classification accuracies differed between the two models (CSP, LSTM) or between the experimental conditions (nomov, mov), a repeated-measures ANOVA was conducted on the accuracy scores of all subjects with preprocessed data from both conditions (n = 18).

For SPoC, in addition to the aforementioned within-participants permutation approach yielding a single *p*-value for each component, group-level statistics were assessed: The hypothesis of a negative correlation between alpha power and emotional arousal was tested with a one-sample, one-tailed *t*-test on the correlation values between z and z_{est}, which assessed whether the mean correlation value per condition was significantly lower than 0.

The code for preprocessing of the data, the three prediction models, and the statistical evaluation is available on GitHub (https://github.com/SHEscher/NeVRo).

Acknowledgements

Thanks to Mina Jamshidi Idaji, Mert Akbal, Nicolas Endres, Alireza Tarikhi, Cade McCall, Firat Sansal, Wojciech Samek, Klaus Robert-Müller, Frederik Harder, Kristof Schütt, Leila Arras and Stefan Haufe.

References

Bishop, C. M. (2007). *Pattern Recognition and Machine Learning* (1st ed. 2006. Corr. 2nd printing 2011). Springer.

Fisher, R. A. (1936). The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, *7*(2), 179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x

Kothe, C., & Makeig, S. (2013). BCILAB: A platform for brain-computer interface development. *Journal of Neural Engineering*. https://doi.org/10.1088/1741-2560/10/5/056014

Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*(2), 365–411. https://doi.org/10.1016/S0047-259X(03)00096-4

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., … Zheng, X. (2015). TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *ArXiv*, 1–19.

Adolphs, R., Mlodinow, L., & Barrett, L. F. (2019). What is an emotion? *Current Biology*, *29*(20), R1060–R1064. https://doi.org/10.1016/j.cub.2019.09.008

Agrawal, P., Stansbury, D., Malik, J., & Gallant, J. L. (2014). Pixels to Voxels: Modeling Visual Representation in the Human Brain. *ArXiv*, 1–15.

Arras, L., Montavon, G., Müller, K.-R., & Samek, W. (2017). Explaining Recurrent Neural Network Predictions in Sentiment Analysis. *ArXiv*.

Bach, D. R., Friston, K. J., & Dolan, R. J. (2010). Analytic measures for quantification of arousal from spontaneous skin conductance fluctuations. *International Journal of Psychophysiology*, *76*(1), 52–55. https://doi.org/10.1016/j.ijpsycho.2010.01.011

Barrett, L. F. (2017). *How Emotions Are Made: The Secret Life of the Brain*. Houghton Mifflin Harcourt.

Bashivan, P., Rish, I., Yeasin, M., & Codella, N. (2016). Learning Representations from EEG with Deep Recurrent-Convolutional Neural Networks. *ICLR*, 1–15.

Berger, H. (1929). Über das Elektrenkephalogramm des Menschen. *Archiv Für Psychiatrie Und Nervenkrankheiten*, *87*, 527–570.

Bergstra, J., & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research 13*, *13*, 281–305.

Berntsen, D., & Rubin, D. C. (2006). Emotion and vantage point in autobiographical. *Cognition and Emotion*, *20*(8), 1193–1215. https://doi.org/10.1080/02699930500371190

Biessmann, F., Gaebler, M., Lamke, J.-P., Ju, U. J., Hetzer, S., Wallraven, C., & Müller, K.-R. (2014). Data-driven Multisubject Neuroimaging Analyses for Naturalistic Stimuli. *IEEE International Workshop on Pattern Recognition in Neuroimaging*, 1–4.

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., & Muller, K. (2008). Optimizing Spatial filters for Robust EEG Single-Trial Analysis. *IEEE Signal Processing Magazine*, *25*(1), 41–56. https://doi.org/10.1109/MSP.2008.4408441

Blascovich, J. (1990). Individual Differences in Physiological Arousal and Perception of Arousal. *Personality and Social Psychology Bulletin*, *16*(4), 665–675.

Bouchard, S., Dumoulin, S., Robillard, G., Guitard, T., Klinger, É., Forget, H., Loranger, C., & Roucaut, F. X. (2017). Virtual reality compared with in vivo exposure in the treatment of social anxiety disorder: A three-arm randomised controlled trial. *The British Journal of Psychiatry*, *210*, 276–283. https://doi.org/10.1192/bjp.bp.116.184234

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), 602–607. https://doi.org/10.1111/j.1469-8986.2008.00654.x

Bridwell, D. A., Cavanagh, J. F., Collins, A. G. E., Nunez, M. D., Srinivasan, R., Stober, S., & Calhoun, V. D. (2018). Moving Beyond ERP Components: A Selective Review of Approaches to Integrate EEG and Behavior. *Frontiers in Human Neuroscience*, *12*. https://doi.org/10.3389/fnhum.2018.00106

Buzsáki, G. (2006). *Rhythms of the Brain*. Oxford University Press. https://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195301069.001.0001/acprof-9780195301069

Cesarei, A. D., & Codispoti, M. (2011). Affective modulation of the LPP and α-ERD during picture viewing. *Psychophysiology*, *48*(10), 1397–1404. https://doi.org/10.1111/j.1469-8986.2011.01204.x

Chaumon, M., Bishop, D. V. M., & Busch, N. A. (2015). A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *Journal of Neuroscience Methods*, *250*, 47–63. https://doi.org/10.1016/j.jneumeth.2015.02.025

Clevert, D.-A., Unterthiner, T., & Hochreiter, S. (2016). Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). *ArXiv:1511.07289 [Cs]*. http://arxiv.org/abs/1511.07289

Cohen, M. X. (2018). Using spatiotemporal source separation to identify prominent features in multichannel data without sinusoidal filters. *European Journal of Neuroscience*, *48*(7), 2454–2465. https://doi.org/10.1111/ejn.13727

Cohen, S. S., Madsen, J., Touchan, G., Robles, D., Lima, S. F. A., Henin, S., & Parra, L. C. (2018). Neural engagement with online educational videos predicts learning performance for individual students. *Neurobiology of Learning and Memory*, *155*, 60–64. https://doi.org/10.1016/j.nlm.2018.06.011

Daehne, S., Meinecke, F. C., Haufe, S., Hoehne, 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. https://doi.org/10.1016/j.neuroimage.2013.07.079

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. https://doi.org/10.1016/j.neuroimage.2013.07.079

Dalgleish, T. (2004). The emotional brain. *Nature Reviews Neuroscience*, *5*(7), 583–589.

https://doi.org/10.1038/nrn1432

de Cheveigné, A., & Parra, L. C. (2014). Joint decorrelation, a versatile tool for multichannel data analysis. *NeuroImage*, *98*, 487–505. https://doi.org/10.1016/j.neuroimage.2014.05.068

Debener, S., Minow, F., Emkes, R., Gandras, K., & de Vos, M. (2012). How about taking a low-cost, small, and wireless EEG for a walk?: EEG to go. *Psychophysiology*, *49*(11), 1617–1621. https://doi.org/10.1111/j.1469-8986.2012.01471.x

Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y., & MÃ¼hlberger, A. (2015). The impact of perception and presence on emotional reactions: A review of research in virtual reality. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00026

Dietterich, T. G. (2000). Ensemble Methods in Machine Learning. *Multiple Classifier Systems*, 1–15. https://doi.org/10.1007/3-540-45014-9_1

Dmochowski, J. P., Sajda, P., Dias, J., & Parra, L. C. (2012). Correlated Components of Ongoing EEG Point to Emotionally Laden Attention – A Possible Marker of Engagement? *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00112

Doetsch, P., Kozielski, M., & Ney, H. (2014). Fast and robust training of recurrent neural networks for offline handwriting recognition. *IEEE Conference on Frontiers in Handwriting Recognition*, 279–284.

Donahue, J., Hendricks, L. A., Rohrbach, M., Venugopalan, S., Guadarrama, S., Saenko, K., & Darrell, T. (2015). Long-term Recurrent Convolutional Networks for Visual Recognition and Description. *IEEE Conference on Computer Vision and Pattern Recognition*, 2625–2634.

Duffy, E. (1957). The psychological significance of the concept of 'arousal' or 'activation'. *The Philosophical Review*, *64*(5), 265–275.

Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology*, *65*(1), 45–55. https://doi.org/10.1037//0022-3514.65.1.45

Gaebler, M., Biessmann, F., Lamke, J.-P., Müller, K.-R., Walter, H., & Hetzer, S. (2014). Stereoscopic depth increases intersubject correlations of brain networks. *NeuroImage*, *100*, 427–434. https://doi.org/10.1016/j.neuroimage.2014.06.008

Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural Networks and the Bias/Variance Dilemma. *Neural Computation*, *4*, 1–58. https://doi.org/10.1162/neco.1992.4.1.1

Gibson, J. J. (1979). *The ecological approach to visual perception*.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. http://www.deeplearningbook.org

Graves, A., Jaitly, N., & Mohamed, A.-R. (2013). Hybrid speech recognition with deep bidirectional LSTM. *Automatic Speech Recognition and Understanding*, 273–278.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(10), 2222–2232.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, *1*(1), 114–125. https://doi.org/10.1007/s40881-015-0004-4

Gross, J. J., & Muñoz, R. F. (1995). Emotion Regulation and Mental Health. *Clinical Psychology: Science and Practice*, *2*(2), 151–164. https://doi.org/10.1111/j.1468-2850.1995.tb00036.x

Gwin, J. T., Gramann, K., Makeig, S., & Ferris, D. P. (2010). Removal of Movement Artifact From High-Density EEG Recorded During Walking and Running. *Journal of Neurophysiology*, *103*(6), 3526–3534. https://doi.org/10.1152/jn.00105.2010

Haegens, S., Cousijn, H., Wallis, G., Harrison, P. J., & Nobre, A. C. (2014). Inter- and intra-individual variability in alpha peak frequency. *NeuroImage*, *92*, 46–55. https://doi.org/10.1016/j.neuroimage.2014.01.049

Haller, M., Donoghue, T., Peterson, E., Varma, P., Sebastian, P., Gao, R., Noto, T., Knight, R. T., Shestyuk, A., & Voytek, B. (2018). *Parameterizing neural power spectra* [Preprint]. Neuroscience. https://doi.org/10.1101/299859

Hasson, U., Nastase, S. A., & Goldstein, A. (2020). Direct Fit to Nature: An Evolutionary Perspective on Biological and Artificial Neural Networks. *Neuron*, *105*(3), 416–434. https://doi.org/10.1016/j.neuron.2019.12.002

Haufe, S., Dähne, S., & Nikulin, V. V. (2014). Dimensionality reduction for the analysis of brain oscillations. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2014.06.073

Haufe, S., & Ewald, A. (2019). A simulation framework for benchmarking EEG-based brain connectivity estimation methodologies. *Brain Topography*, 32(4), 625-642

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, *87*, 96–110. https://doi.org/10.1016/j.neuroimage.2013.10.067

Hefron, R. G., Borghetti, B. J., Christensen, J. C., & Schubert, C. M. (2017). Deep long short-term memory structures model temporal dependencies improving cognitive workload estimation. *Pattern Recognition Letters*, *94*, 96–104. https://doi.org/10.1016/j.patrec.2017.05.020

Herman, P., Prasad, G., McGinnity, T. M., & Coyle, D. (2008). Comparative Analysis of Spectral Approaches to Feature Extraction for EEG-Based Motor Imagery Classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *16*(4), 317–326.

Hestness, J., Narang, S., Ardalani, N., Diamos, G., Jun, H., Kianinejad, H., Patwary, M. M. A., Yang, Y., & Zhou, Y. (2017). Deep Learning Scaling is Predictable, Empirically. *ArXiv:1712.00409 [Cs, Stat]*. http://arxiv.org/abs/1712.00409

Hochreiter, S., & Schmidhuber, J. (1995). *Long Short Term Memory* (No. 1993; pp. 1–8).

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780.

Huang, Y., Parra, L. C., & Haufe, S. (2016). The New York Head—A precise standardized volume conductor model for EEG source localization and tES targeting. *NeuroImage*,

140, 150–162. https://doi.org/10.1016/j.neuroimage.2015.12.019

Huk, A., Bonnen, K., & He, B. J. (2018). Beyond Trial-Based Paradigms: Continuous Behavior, Ongoing Neural Activity, and Natural Stimuli. *The Journal of Neuroscience*, *38*(35), 7551–7558. https://doi.org/10.1523/JNEUROSCI.1920-17.2018

Idaji, M. J., Müller, K. R., Nolte, G., Maess, B., Villringer, A., & Nikulin, V. V. (2020). Nonlinear interaction decomposition (NID): A method for separation of cross-frequency coupled sources in human brain. *NeuroImage*, *211*, 116599. https://doi.org/10.1016/j.neuroimage.2020.116599

James, W. (1884). What is an Emotion? *Mind*, *9*(34), 188–205.

James, W. (1890). *The Pinciples of Psychology in Two Volumes*. Macmillan.

Jensen, O., & Mazaheri, A. (2010). Shaping Functional Architecture by Oscillatory Alpha Activity: Gating by Inhibition. *Frontiers in Human Neuroscience*, *4*. https://doi.org/10.3389/fnhum.2010.00186

Kaplan, R. L., Levine, L. J., Lench, H. C., & Safer, M. A. (2016). Forgetting feelings: Opposite biases in reports of the intensity of past emotion and mood. *Emotion*. https://doi.org/10.1037/emo0000127

Kelly, S. P., Lalor, E. C., Reilly, R. B., & Foxe, J. J. (2006). Increases in Alpha Oscillatory Power Reflect an Active Retinotopic Mechanism for Distracter Suppression During Sustained Visuospatial Attention. *Journal of Neurophysiology*, *95*(6), 3844–3851. https://doi.org/10.1152/jn.01234.2005

Khaligh-Razavi, S.-M., & Kriegeskorte, N. (2014). Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation. *PLOS Computational Biology*, *10*(11), e1003915. https://doi.org/10.1371/journal.pcbi.1003915

Kingma, D. P., & Ba, J. L. (2015). Adam: A Method for Stochastic Optimization. *ICLR*, 1–15.

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*(12), 606–617. https://doi.org/10.1016/j.tics.2012.10.007

Koelstra, S., Muhl, C., Soleymani, M., Jong-Seok Lee, 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

Kragel, P. A., & LaBar, K. S. (2013). Multivariate pattern classification reveals autonomic and experiential representations of discrete emotions. *Emotion*, *13*(4), 681–690. https://doi.org/10.1037/a0031820

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, *84*(3), 394–421. https://doi.org/10.1016/j.biopsycho.2010.03.010

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540. https://doi.org/10.1038/nn.2303

Krusienski, D. J., Grosse-Wentrup, M., Galán, F., Coyle, D., Miller, K. J., Forney, E., & Anderson,

C. W. (2011). Critical issues in state-of-the-art brain–computer interface signal processing. *Journal of Neural Engineering*, *8*(2). https://doi.org/10.1088/1741-2560/8/2/025002.Critical

Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010). Feelings change: Accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology*, *99*(6), 1042–1060. https://doi.org/10.1037/a0020962

Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological Bulletin*, *139*(4), 917–940. https://doi.org/10.1037/a0030811

Lapuschkin, S. (2019). *Opening the machine learning black box with Layer-wise Relevance Propagation*. http://dx.doi.org/10.14279/depositonce-7942

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, T. W., Girolami, M., & Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Computation*, *11*(2), 417–441. https://doi.org/10.1162/089976699300016719

Lettieri, G., Handjaras, G., Ricciardi, E., Leo, A., Papale, P., Betta, M., Pietrini, P., & Cecchetti, L. (2019). Emotionotopy in the human right temporo-parietal cortex. *Nature Communications*, *10*(1), 5568. https://doi.org/10.1038/s41467-019-13599-z

Levine, L. J., & Safer, M. A. (2002). Sources of Bias in Memory for Emotions. *Current Directions in Psychological Science*, *11*(5), 169–173. https://doi.org/10.1111/1467-8721.00193

Lindquist, K. A. (2013). Emotions Emerge from More Basic Psychological Ingredients: A Modern Psychological Constructionist Model. *Emotion Review*, *5*(4), 356–368. https://doi.org/10.1177/1754073913489750

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences*, *35*(03), 121–143. https://doi.org/10.1017/S0140525X11000446

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., & Yger, F. (2018). A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update. *Journal of Neural Engineering*, *15*(3), 031005. https://doi.org/10.1088/1741-2552/aab2f2

Luft, C. D. B., & Bhattacharya, J. (2015). Aroused with heart: Modulation of heartbeat evoked potential by arousal induction and its oscillatory correlates. *Scientific Reports*, *5*(1). https://doi.org/10.1038/srep15717

Luong, M., Sutskever, I., Le, Q. V., Vinyals, O., & Zaremba, W. (2015). Addressing the Rare Word Problem in Neural Machine Translation. *ArXiv*.

Marcotti, P., & Jacques, P. L. S. (2018). Shifting visual perspective during memory retrieval reduces the accuracy of subsequent memories. *Memory*. https://doi.org/10.1080/09658211.2017.1329441

Matusz, P. J., Dikker, S., Huth, A. G., & Perrodin, C. (2019). Are We Ready for Real-world Neuroscience? *Journal of Cognitive Neuroscience*, *31*(3), 327–338. https://doi.org/10.1162/jocn_e_01276

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition & Emotion*, *23*(2), 209–237. https://doi.org/10.1080/02699930802204677

McCall, C., Hildebrandt, L. K., Bornemann, B., & Singer, T. (2015). Physiophenomenology in retrospect: Memory reliably reflects physiological arousal during a prior threatening experience. *Consciousness and Cognition*, *38*, 60–70. https://doi.org/10.1016/j.concog.2015.09.011

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*. https://doi.org/10.3389/fnhum.2016.00170

Mierau, A., Klimesch, W., & Lefebvre, J. (2017). State-dependent alpha peak frequency shifts: Experimental evidence, potential mechanisms and functional implications. *Neuroscience*, *360*, 146–154. https://doi.org/10.1016/j.neuroscience.2017.07.037

Mikutta, C., Altorfer, A., Strik, W., & Koenig, T. (2012). Emotions, Arousal, and Frontal Alpha Rhythm Asymmetry During Beethoven's 5th Symphony. *Brain Topography*, *25*(4), 423–430. https://doi.org/10.1007/s10548-012-0227-0

Moosmann, M., Ritter, P., Krastel, I., Brink, A., Thees, S., Blankenburg, F., Taskin, B., Obrig, H., & Villringer, A. (2003). Correlates of alpha rhythm in functional magnetic resonance imaging and near infrared spectroscopy. *NeuroImage*, *20*(1), 145–158. https://doi.org/10.1016/S1053-8119(03)00344-6

Naumann, L., Schultze-Kraft, M., Sven, D., & Blankertz, B. (2016). Prediction of Difficulty Levels in Video Games from Ongoing EEG. *International Workshop on Symbiotic Interaction*, 125–136. https://doi.org/10.1007/978-3-319-57753-1

Neal, B., Mittal, S., Baratin, A., Tantia, V., Scicluna, M., Lacoste-Julien, S., & Mitliagkas, I. (2019). A Modern Take on the Bias-Variance Tradeoff in Neural Networks. *ArXiv:1810.08591 [Cs, Stat]*. http://arxiv.org/abs/1810.08591

Nikulin, V. V., Nolte, G., & Curio, G. (2011). A novel method for reliable and fast extraction of neuronal EEG/MEG oscillations on the basis of spatio-spectral decomposition. *NeuroImage*, *55*(4), 1528–1535. https://doi.org/10.1016/j.neuroimage.2011.01.057

Nurse, E. S., Karoly, P. J., Grayden, D. B., & Freestone, D. R. (2015). A Generalizable Brain-Computer Interface (BCI) Using Machine Learning for Feature Discovery. *PLOS ONE*, *10*(6), e0131328. https://doi.org/10.1371/journal.pone.0131328

Olbrich, S., Mulert, C., Karch, S., Trenner, M., Leicht, G., Pogarell, O., & Hegerl, U. (2009). EEG-vigilance and BOLD effect during simultaneous EEG/fMRI measurement. *NeuroImage*, *45*(2), 319–332. https://doi.org/10.1016/j.neuroimage.2008.11.014

Opitz, D., & Maclin, R. (1999). Popular Ensemble Methods: An Empirical Study. *Journal of Artificial Intelligence Research*, *11*, 169–198. https://doi.org/10.1613/jair.614

Pan, X., & Hamilton, A. F. de C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal*

of Psychology (London, England: 1953), *109*(3), 395–417. https://doi.org/10.1111/bjop.12290

Parra, L. C., Spence, C. D., Gerson, A. D., & Sajda, P. (2005). Recipes for the linear analysis of EEG. *NeuroImage*, *28*(2), 326–341. https://doi.org/10.1016/j.neuroimage.2005.05.032

Pascual-Marqui, R. D. (2007). Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: Exact, zero error localization. *ArXiv:0710.3341 [Math-Ph, Physics:Physics, q-Bio]*. http://arxiv.org/abs/0710.3341

Pfaff, D. W., Martin, E. M., & Faber, D. (2012). Origins of arousal: Roles for medullary reticular neurons. *Trends in Neurosciences*, *35*(8), 468–476. https://doi.org/10.1016/j.tins.2012.04.008

Poulsen, A. T., Kamronn, S., Dmochowski, J., Parra, L. C., & Hansen, L. K. (2016). EEG in the classroom: Synchronised neural recordings during video presentation. *ArXiv*, 1–14.

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. https://doi.org/10.1109/86.895946

Rebenitsch, L., & Owen, C. (2016). Review on cybersickness in applications and visual displays. *Virtual Reality*, *20*(2), 101–125. https://doi.org/10.1007/s10055-016-0285-9

Rivet*, B., Souloumiac, A., Attina, V., & Gibert, G. (2009). xDAWN Algorithm to Enhance Evoked Potentials: Application to Brain–Computer Interface. *IEEE Transactions on Biomedical Engineering*, *56*(8), 2035–2043. https://doi.org/10.1109/TBME.2009.2012869

Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review*, *33*(1), 1–39. https://doi.org/10.1007/s10462-009-9124-7

Ruder, S. (2017). An overview of gradient descent optimization. *ArXiv*, 1–14.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Russell, J. A., & Feldman Barrett, L. (1999). Core Affect, Prototypical Emotional Episodes, and Other Things Called Emotion: Dissecting the Elephant. *Journal of Personality and Social Psychology*, *76*(5), 805–819. https://doi.org/10.1037/0022-3514.76.5.805

Saarimäki, H., Gotsopoulos, A., Jääskeläinen, I. P., Lampinen, J., Vuilleumier, P., Hari, R., Sams, M., & Nummenmaa, L. (2016). Discrete Neural Signatures of Basic Emotions. *Cerebral Cortex*, *26*(6), 2563–2573. https://doi.org/10.1093/cercor/bhv086

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*, 116893. https://doi.org/10.1016/j.neuroimage.2020.116893

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Schmidt, R. E., Gay, P., D'Acremont, M., & van der Linden, M. (2008). A German adaptation of the UPPS Impulsive Behavior Scale: Psychometric properties and factor structure. *Swiss Journal of Psychology*, *67*, 107–112.

Schubring, D., & Schupp, H. T. (2019). Affective picture processing: Alpha- and lower beta-band desynchronization reflects emotional arousal. *Psychophysiology*, *56*(8), e13386. https://doi.org/10.1111/psyp.13386

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, *17*(11), 565–573. https://doi.org/10.1016/j.tics.2013.09.007

Shamay-Tsoory, S. G., & Mendelsohn, A. (2019). Real-Life Neuroscience: An Ecological Approach to Brain and Behavior Research. *Perspectives on Psychological Science*, *14*(5), 841–859. https://doi.org/10.1177/1745691619856350

Sharbrough, F., Chatria, G.-E., Lesser, R., Lüders, H., Nuwer, M., & Picton, T. (1991). American Electroencephalographic Society Guidelines for Standard Electrode Position Nomenclature. *Journal of Clinical Neurophysiology*, *8*(2), 200–202.

Sharples, S., Cobb, S., Moody, A., & Wilson, J. R. (2008). Virtual reality induced symptoms and effects (VRISE): Comparison of head mounted display (HMD), desktop and projection display systems. *Displays*, *29*(2), 58–69. https://doi.org/10.1016/j.displa.2007.09.005

Siegel, E. H., Sands, M. K., Van den Noortgate, W., Condon, P., Chang, Y., Dy, J., Quigley, K. S., & Barrett, L. F. (2018). Emotion fingerprints or emotion populations? A meta-analytic investigation of autonomic features of emotion categories. *Psychological Bulletin*, *144*(4), 343–393. https://doi.org/10.1037/bul0000128

Spielberger, C. D. (1983). *Manual for the State–Trait Anxiety Inventory (Form Y)*. Mind Garden.

Spielberger, C. D. (1989). *State–Trait Anxiety Inventory: A comprehensive bibliography*. Consulting Psychologists Press.

Sturm, I., Lapuschkin, S., Samek, W., & Müller, K.-R. (2016). Interpretable deep neural networks for single-trial EEG classification. *Journal of Neuroscience Methods*, *274*, 141–145. https://doi.org/10.1016/j.jneumeth.2016.10.008

Uusberg, A., Uibo, H., Kreegipuu, K., & Allik, J. (2013). EEG alpha and cortical inhibition in affective attention. *International Journal of Psychophysiology*, *89*(1), 26–36. https://doi.org/10.1016/j.ijpsycho.2013.04.020

Van Diepen, R. M., Foxe, J. J., & Mazaheri, A. (2019). The functional role of alpha-band activity in attentional processing: The current zeitgeist and future outlook. *Current Opinion in Psychology*, *29*, 229–238. https://doi.org/10.1016/j.copsyc.2019.03.015

Vasser, M., & Aru, J. (2020). Guidelines for Immersive Virtual Reality in Psychological Research. *Current Opinion in Psychology*, S2352250X20300683. https://doi.org/10.1016/j.copsyc.2020.04.010

Weech, S., Kenny, S., & Barnett-Cowan, M. (2019). Presence and Cybersickness in Virtual Reality Are Negatively Related: A Review. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00158

Whiteside, S. P., & Lynam, D. R. (2001). The five factor model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences*, *30*(4), 669–689.

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. https://doi.org/10.1177/0956797612464242

Wöllmer, M., Eyben, F., Reiter, S., Cox, C., Douglas-Cowie, E., & Cowie, R. (2008). Abandoning Emotion Classes—Towards Continuous Emotion Recognition with Modelling of Long-Range Dependencies. *Interspeech*, 597–600.

Wöllmer, M., Metallinou, A., Eyben, F., Schuller, B., & Narayanan, S. (2010). Context-Sensitive Multimodal Emotion Recognition from Speech and Facial Expression using Bidirectional LSTM Modeling. *Interspeech*, 2362–2365.

Wundt, W. M. (1897). *Outlines of Psychology*. Engelmann.

Zaremba, W., Sutskever, I., & Vinyals, O. (2015). Recurrent Neural Network Regularization. *ICLR*, 1–8.

Zuure, M. B., & Cohen, M. X. (2020). Narrowband multivariate source separation for semi-blind discovery of experiment contrasts. *BioRxiv*, 2020.03.09.983635. https://doi.org/10.1101/2020.03.09.983635