

# Emotion Recognition from EEG During Self-Paced Emotional Imagery

Christian Andreas Kothe and Scott Makeig  
Swartz Center for Computational Neuroscience  
University of California San Diego  
La Jolla, CA, USA  
{christian,scott}@scn.ucsd.edu

Julie Anne Onton  
Naval Health Research Center  
San Diego, CA, USA  
julie.onton@med.navy.mil

**Abstract**—Here we present an analysis of a 12-subject electroencephalographic (EEG) data set in which participants were asked to engage in prolonged, self-paced episodes of guided emotion imagination with eyes closed. Our goal is to correctly predict, given a short EEG segment, whether the participant was imagining a positive respectively negative-valence emotional scenario during the given segment using a predictive model learned via machine learning. The challenge lies in generalizing to novel (i.e., previously unseen) emotion episodes from a wide variety of scenarios including love, awe, frustration, anger, etc. based purely on spontaneous oscillatory EEG activity without stimulus event-locked responses. Using a variant of the Filter-Bank Common Spatial Pattern algorithm, we achieve an average accuracy of 71.3% correct classification of binary valence rating across 12 different emotional imagery scenarios under rigorous block-wise cross-validation.

**Keywords**—*emotion; valence; brain-computer interface; EEG; machine learning; guided imagery*

## I. INTRODUCTION

Emotion recognition is a central topic in affective computing [1] that has received increasing attention over the past decade. Nearly all recent studies of emotion recognition have adopted a machine learning approach in which a predictive model, typically a classifier, is trained on features of a set of biophysiological data and then tested on a separate data set (usually obtained from the same participant and/or session). Perhaps the most prolific branch of the field deals with emotion recognition from multimodal data including heart rate, galvanic skin response (GSR), Electromyography (EMG), Electrooculography (EOG), face video, and/or pupil dilation, and is therefore concerned with the extraction and fusion of multimodal features into a single predictor [2,3]. In contrast, we are here concerned with emotion recognition purely from EEG signals, which has itself spurred considerable interest in recent years [4].

Most EEG-based emotion recognition studies employ external stimulus presentations to elicit emotional states in their participants using emotion-laden pictures (e.g., utilizing the International Affective Picture System [5,6]), sounds [7],

movie clips [8,9], or music [10]. In contrast, we are here concerned with inwardly imagined and felt emotions elicited by the participants' own imagination or recall of emotionally loaded memories, an experiment design and data set reported in [11]. While a small group of studies have applied a recall paradigm to elicit emotions in the context of machine learning [12], the recall duration has been short (up to a few seconds) and/or did not allow the subject to proceed at his or her own pace, factors that may limit the attainable depth and focus of their emotional experience. To bypass these limitations, this experiment used a guided imagery paradigm [13] in which participants are first invited to enter into to a deeply relaxed state via a pre-recorded narrative, and are then invited to imagine experiencing a series of emotion-laden scenarios, separated by guided relaxations. Participants sat in a comfortable chair with eyes closed and were encouraged to exercise their emotional imagination for as long as they could, pressing a handheld button first when they began to somatically experience the suggested emotion, and again when this experience waned or they were otherwise ready to continue (in practice, after 1-5 min).

The unique characteristics of this experiment design pose several analysis challenges. In particular, there were no condition repetitions; each of the 15 imagined scenarios was unique. This forces us to adopt a conservative cross-validation approach, leaving out complete blocks rather than more usual leave-one-trial-out or randomized cross-validation approaches. Furthermore, in our study we test our classifiers on previously unseen conditions (for example, testing an emotion valence classifier trained during guided imagery of love on data from guided imagery of awe). As a side effect, this analysis may yield some of the strongest so-far presented evidence for or against the practicality of general-purpose recognition of emotional valence from EEG.

## II. EXPERIMENTAL TASK

The experiment was performed in a dimly-lit room equipped with a comfortable chair; the experimenter was seated in a separate control room. The recording session (ca. 80 min on average) was performed with eyes closed. Instructions

were delivered in the form of pre-recorded verbal narratives via earbud speakers. The recording session began with a 2-min resting period followed by a verbal explanation of the task and a verbally guided relaxation exercise of about 5 min to promote a relaxed, inwardly-focused state. The subsequent main task was a sequence of 15 blocks, each beginning with a (15-30 sec) guided imagination narrative describing a particular emotion, followed by an imagination period that lasted until the participant pressed the response button a second time (on average, after  $218 \pm \text{s.d. } 94$  sec), and each ending with a 15-sec relaxation narrative to restore a neutral state. Each induction narrative began with a short description of the emotion followed by suggestions of one or more circumstances in which the target emotion might be vividly experienced. Participants were instructed to use whatever imagery they deemed suitable for stimulating a vivid and embodied experience of the suggested emotion, and were encouraged to pay attention to somatic sensations associated with the target emotion. They were asked to take as much time as they needed to recall or imagine a scenario that would induce a realistic experience of the described emotion. They were encouraged to experience each emotion for 3–5 min, pressing a left hand-held button when their emotional experience began and then again when it subsided or they were ready to move on. To minimize participant fatigue, the emotion sequence was chosen to alternate pseudo-randomly between 8 selected positive-valence emotions (love, joy, happiness, relief, compassion, contentment, excitement, awe) and 7 negative-valence emotions (anger, jealousy, disgust, frustration, fear, sadness, grief). The experiment ended after another 2-min silent resting period. After the experiment, all participants stated they felt that they had experienced realistic emotional states using the verbal narratives and their own imagination.

### III. DATA ACQUISITION

Sixteen young adult volunteers (mixed gender,  $25.5 \pm \text{s.d. } 5$  years) participated under informed consent in accordance with University of California San Diego institutional review board requirements. We here restrict ourselves to 12 subjects since four recordings had partially missing marker information. The study included 16 further participants, not analyzed here, in a modified experiment protocol (see [11]).

EEG data were collected from 250 gel-based scalp electrodes, plus four infraocular and two electrocardiographic (ECG) placements using a BioSemi ActiveTwo (Biosemi, NL) amplifier with 24-bit resolution. Caps with a custom whole-head montage that covered most of the skull, forehead, and lateral face surface were used, omitting chin and fleshy cheek areas. Locations of the electrodes and skull landmarks for each participant were digitized (Polhemus). To expedite computations in the subsequent analysis the data were here resampled to 128 Hz and reduced to a subset of 124 evenly-distributed EEG channels. Since here we attempt to classify positive versus negative valence, we excluded three emotion conditions (excitement, disgust, and compassion) that did not fall clearly into the positive or negative valence categories. Further, we discarded the first two remaining blocks (during which we assumed the participant was still adapting to the task), leaving 10 blocks per participant, on average 5 labeled as

positive and 5 as negative valence both by participants and in separate group ratings.

## IV. METHOD

To predict the valence of the emotion experienced by the participant from single EEG segments, we employ a predictive model that estimates, from a given short EEG segment (here 6 sec.), the probability that the subject experienced a positive or negative valence emotion during that period — a binary classification. Our approach relies on changes in the power spectrum of short-time stationary oscillatory EEG processes within standard EEG frequency bands (delta, theta, alpha, beta, low gamma) at unknown source locations. We employ a variant of the Filter-Bank Common Spatial Pattern method (FBCSP [14]) that finds optimal spatial filters in whose outputs power spectral differences are maximally discriminative between conditions. Here, we use a sparse feature-selecting classifier (logistic regression with elastic-net regularization [15]) instead of performing feature selection in a separate step. The method, implemented using the open-source BCILAB toolbox [16], can be trained in under 5 min on a recent PC and can be used to perform real-time classification.

### A. Data Pre-Processing and Segmentation

The continuous EEG data were first high-pass filtered using a Butterworth IIR filter with a 0.1-1 Hz transition band. Then,  $N=40$  equally spaced segments  $X_i$ , each 6 sec in length, were extracted from each block (giving on average about 50% overlap for successive segments). We excluded the first 60 sec and last 10 sec of each emotion imagination block to focus on the period of maximum engagement. Each extracted segment was then associated with a label  $y_i \in \{+1, -1\}$  corresponding to the positive or negative valence of the block label.

### B. Single-Trial Classification

Given a short high-pass filtered EEG segment  $X \in \mathbf{R}^{C \times T}$ , the data are first band-pass filtered to  $n_f=5$  separate pre-defined EEG frequency bands (delta, 0.5-3 Hz; theta, 4-7 Hz; alpha, 8-12 Hz; beta, 13-30 Hz; gamma, 31-42 Hz). For simplicity filtering is applied using temporal filters  $B_f$  comprising FFT, spectral weighting, and inverse FFT transforms. Each resulting segment of band-pass filtered EEG is further linearly spatially filtered by a matrix  $W_f = [w_{1,f}, w_{2,f}, \dots, w_{k,f}]^T$  of  $k=8$  band-specific spatial filters, which gives an 8-channel signal whose per-channel log-variances are taken as the feature vector  $x_f$  for frequency band  $f$  as

$$x_f = \log(\text{Var}[W_f X B_f])$$

The resulting feature vectors  $x_f$ , concatenated into a single feature vector for trial  $\mathbf{x}=[x_1, x_2, \dots, x_f]$ , are then mapped onto a trial probability value using a generalized linear model (GLM) with a logistic link function

$$P(Y = y|\mathbf{x}) = \frac{1}{1 + e^{-y(b + \theta^T \mathbf{x})}}$$

This value represents the probability that the given EEG segment is of positive (resp. negative) class, here indexing the valence of the emotion being experienced by the participant.

### C. Model Calibration

To calculate the spatial filters  $\mathbf{W}_f$  for a given frequency band  $f$  from a collection of high-pass filtered training trial segments  $\mathbf{X}_i$  and associated labels  $y_i$  we first apply the filter  $\mathbf{B}_f$  to each segment and concatenate all filtered segments with positive label into matrix  $\mathbf{X}^{(+)}$  and all filtered negative-labeled segments into matrix  $\mathbf{X}^{(-)}$  to calculate the per-condition covariance matrices  $\mathbf{C}^{(+)} = \mathbf{X}^{(+)}\mathbf{X}^{(+)\top}$  and  $\mathbf{C}^{(-)} = \mathbf{X}^{(-)}\mathbf{X}^{(-)\top}$ , respectively. We then solve the generalized eigenvalue problem

$$\mathbf{C}^{(+)}\mathbf{v} = \lambda\mathbf{C}^{(-)}\mathbf{v}$$

as in the Common Spatial Pattern (CSP) [17] algorithm, giving a matrix of eigenvalues  $\lambda$  and eigenvectors  $\mathbf{V}$  of which we retain the ( $k = 4$ ) components at the upper and lower ends of the eigenvalue spectrum. These vectors are concatenated into the filter matrix  $\mathbf{W}_f$ .

To learn the parameters  $(\boldsymbol{\theta}, b)$  of the generalized linear prediction function, we employ logistic regression with elastic-net regularization (elastic mixing parameter  $\alpha$  fixed at  $\frac{1}{4}$ ). This amounts to solving the convex optimization problem

$$\min_{\boldsymbol{\theta}, b} \frac{1}{N} \sum_t \log(1 + e^{-y_t(\boldsymbol{\theta}^\top \mathbf{x}_t + b)}) + \mu\alpha\|\boldsymbol{\theta}\|_1 + \mu(1 - \alpha)\|\boldsymbol{\theta}\|_2^2$$

for centered and standardized feature vectors  $\mathbf{x}_i$  and their associated labels  $y_i$  extracted from a set of training blocks as described in Section IV(B). To determine the free parameter  $\mu$  the problem is solved repeatedly for a series of candidate values of  $\mu$  and the best value is selected using a leave-one-block-out cross-validation on the training set. This problem can be solved efficiently using the *glmnet* package for MATLAB (The Mathworks, Natick, MA).

The elastic-net penalty was chosen under the assumption that not all frequency bands or spatial filters are relevant (implying sparsity), while at the same time features for neighboring frequencies are likely correlated, suggesting the use of the additional  $l_2$  term to encourage equal weighting for similarly-relevant features.

### D. Performance Evaluation

To assess whether emotional valence can be predicted from single EEG trials, we evaluated the test-set accuracy of the method using a 5-fold block-wise cross-validation on the 10 blocks from each participant. Thus, in each fold of the cross-validation, two successive blocks were declared the test set and model calibration was performed on the 8 remaining training blocks. The resulting predictive model was then tested on the data segments of the two test-set blocks. The classification accuracy of the method was quantified as percent correct given the class labels of the test blocks.

## V. RESULTS

The mean accuracy of our method, across the 12 participants, was 71.3%  $\pm$  14.9%, which is highly significant given the chance level of 50% ( $p < 0.01$  in a standard t-test). Accuracies for individual subjects are depicted in Fig. 1,

although the within-subject sample size is low due to the block design. The relevant spatial patterns (spatial filter inverses obtained from the  $k$  upper and lower rows of  $\mathbf{V}^{-1}$ ) of the FBCSP models for two typical participants are shown in Figs. 2 and 3, respectively, where the classifier in Fig. 2 (#11, 72% accuracy) appears to involve high-frequency scalp muscle activity, while that in Fig. 3 (#9, 69% accuracy) appears to primarily involve lower frequency activity from cortical sources. The displayed patterns quantify the forward projection patterns of the latent sources extracted by the respective spatial filters. Note, in particular, the occurrence of peri-auricular muscle sources with virtually no far-field projection across the scalp (Fig. 2), and dual-symmetric scalp projection patterns in the alpha band, compatible with occipital and parietal brain generators.

## VI. DISCUSSION

The key result of this analysis is, first, that emotion valence can be classified in this task at better-than-chance level, and second that the level of accuracy may be considered almost practical for real-time neurofeedback or emotion-based interface control. These results hold up under block-wise evaluation with clear separation of training and test data, an approach considerably more rigorous than a randomized cross-validation over segments. Furthermore, since the emotional scenarios used in the test blocks were distinct from those in the training blocks (for example, awe vs. love), our results quantify to what extent the learned classifiers could generalize to unseen conditions. This implies that the observed EEG responses share some commonalities across different emotional scenarios of a given valence level (while exhibiting some differences between levels).

Several of the learned spatial filters seem clearly focused on neck and temporal scalp muscle (or EMG) activity in higher frequency bands. Eye movement cues, when relevant, were relevant at low frequencies while, as expected, near 10-Hz alpha frequency band variance was relevant for occipital and parietal brain sources. For some participants, other brain sources were relevant at theta and beta bands, although their locations did not appear to be consistent.

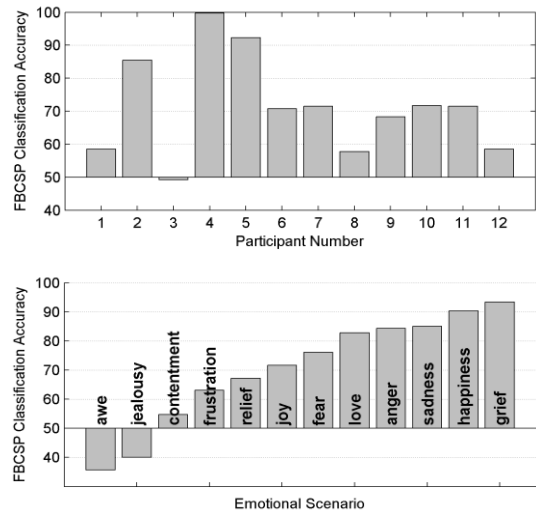


Figure 1. Cross-validated valence classification accuracy across all participants (top) and emotional scenarios (bottom). Chance level is 50%.

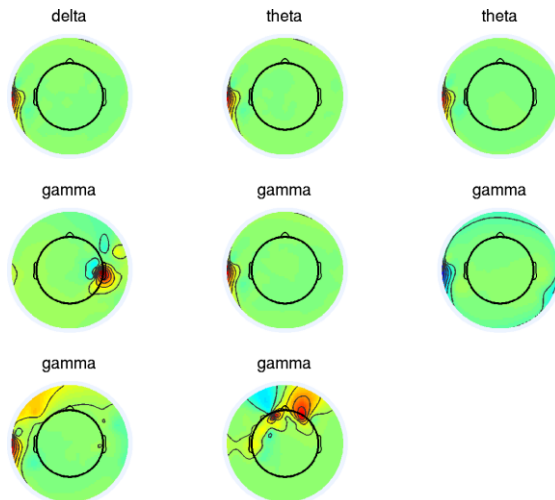


Figure 2. Forward scalp projections of relevant source mixtures selected by the classifier for a participant (#11) for whom the classifier is dominated by scalp muscle activity. The relevant frequency bands are indicated.

## VII. CONCLUSION

We have presented a classifier for experienced emotion valence in an experiment designed to elicit strong emotional experiences. A rigorous analysis of these data produced strong evidence that a key aspect of emotional state – emotional valence – can be classified from a few seconds of spontaneous EEG data. Because of the inclusion of informative muscular activity in the EEG data for most subjects (possibly related to “tensing up” during some emotion states), we cannot conclusively determine from these results to what extent brain-source EEG activity alone would allow for emotion classification (though this should be possible with further analysis of these results). We can state that the extracted features do generalize for these data to novel emotional scenarios, the key requirement for practical usability of emotion recognition systems. At the core of our analysis is an easy-to-implement method to learn and classify informative spectral power changes across standard frequency bands that may be useful for further investigation of emotion recognition from ongoing EEG data.

## REFERENCES

- [1] R. W. Picard, *Affective Computing*, The MIT Press, 1997.
- [2] S. Koelstra *et al.*, “Single trial classification of EEG and peripheral physiological signals for recognition of emotions induced by music videos,” *Brain Informatics*, Springer, 2010, pp. 89–100.
- [3] H. Hamdi, P. Richard, A. Suteau, and P. Allain, “Emotion assessment for affective computing based on physiological responses,” in *IEEE World Congress on Computational Intelligence*, 2012, pp. 1–8.
- [4] M.-N. Kim, M. Kim, E. Oh, and S.-P. Kim, “A review on the computational methods for emotional state estimation from the human EEG,” *Computational and Mathematical Methods in Medicine* vol. 2013, 2013.
- [5] R. Horlings, D. Datcu, and L. J. M. Rothkrantz, “Emotion recognition using brain activity,” in *CompSysTech’08*, 2008, pp. 6.

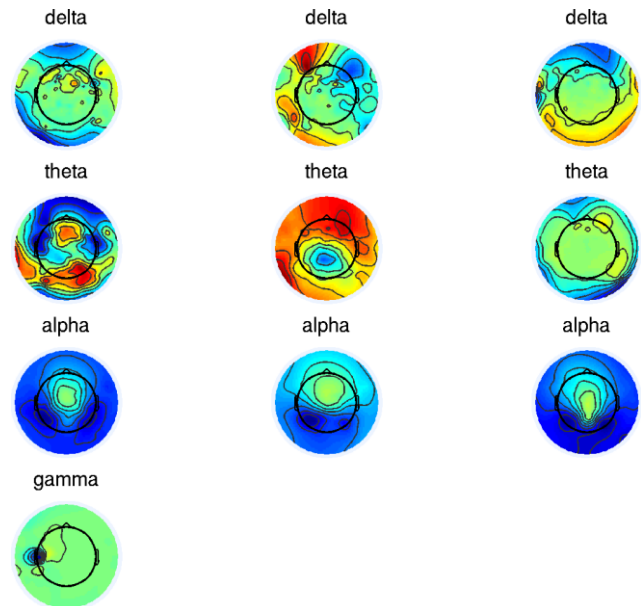


Figure 3. Forward scalp projections of relevant source mixtures selected by the respective classifier for another participant (#9), including both cortical and scalp muscle sources. Other details as in Fig. 2.

- [6] P. D. Bamidis *et al.*, “An integrated approach to emotion recognition for advanced emotional intelligence,” in *Human-Computer Interaction*, Springer Heidelberg, 2009, pp., 565–574.
- [7] Y. Liu, O. Sourina, and M. K. Nguyen, “Real-time EEG-based human emotion recognition and visualization,” in *Intl. Conf. on Cyberworlds*, 2010, pp. 262–269.
- [8] D. Nie, X.-W. Wang, L.-C. Shi, and B.-L. Lu, “EEG-based emotion recognition during watching movies,” in *5th Intl. IEEE/EMBS Conf. on Neural Engineering*, 2011, pp. 667–670.
- [9] M. Murugappan, M. Rizon, R. Nagarajan, and S. Yacoob, “Inferring of human emotional states using multichannel EEG,” *European J. of Scientific Research*, vol. 48(2), 2010, pp. 281–299.
- [10] Y. P. Lin *et al.*, “EEG-based emotion recognition in music listening,” *IEEE Trans. Biomed. Eng.*, vol. 57(7), 2010, pp. 1798–1806.
- [11] J. Onton and S. Makeig, “High-frequency broadband modulations of electroencephalographic spectra,” *Frontiers in Human Neuroscience* vol. 3, 2009.
- [12] G. Chanel, J. Kierkels, M. Soleymani, and T. Pun, “Short-term emotion assessment in a recall paradigm,” *Intl. J. Human-Computer Studies*, vol. 67(8), 2009, pp. 607–627.
- [13] H. L. Bonny, *Music & consciousness: the evolution of guided imagery and music*, Barcelona Publishers, 2002.
- [14] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, “Filter bank common spatial pattern (FBCSP) in brain-computer interface.” In *IEEE Intl. Joint Conf. on Neural Networks*, 2008, pp. 2390–2397.
- [15] H. Zou and T. Hastie, “Regularization and variable selection via the elastic net.” *J. of the Royal Statistical Society: Series B (Statistical Methodology)* vol 67(2), 2005, pp. 301–320.
- [16] A. Delorme *et al.*, “EEGLAB, SIFT, NIFT, BCILAB, and ERICA: new tools for advanced EEG processing,” *Computational Intelligence and Neuroscience*, vol. 10, 2011.
- [17] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, “Optimal spatial filtering of single trial EEG during imagined hand movement,” *IEEE Trans. Rehab. Eng.*, vol. 8(4), 2000, pp. 441–446.