What I Feel and What I Say: Decoding Neurophysiological Correlates of Cognitive and Affective States

Katharina Lingelbach[1,2], Sabrina Gado[1], Jochem W. Rieger[2], Mathias Vukelić[1]
[1]Fraunhofer Institute for Industrial Engineering IAO, Stuttgart, Germany
[2]Department of Psychology, University of Oldenburg, Oldenburg, Germany

Motivation and Aim: For neuroergonomic applications, robust decoding of activation patterns indicating current affective or cognitive states is a crucial step to develop adaptive and collaborative human-machine systems leading to increased performance, safety, and user experience [1-6]. So far, most studies attempting to estimate affective and cognitive states from electroencephalographic (EEG) signals used strictly controlled stimuli and tasks to induce a single, isolated mental state. Classification results by machine learning models were promising when estimating workload but rather modest to poor when estimating affective-emotional states [4]. However, in naturalistic environments, we are seldom confronted with isolated stimuli resulting in one mental state but rather experience a composition of interwoven affective and cognitive states in response to complex stimuli. Especially in naturalistic applications, choosing a good approximation to the ground truth that adequately represents a person's true mental states is decisive step when training decoding models. The ground truth can either be estimated based on characteristics of the experimental condition (i.e., hypothetically estimated) or by asking participants to provide labels based on their experiences via post-hoc ratings or questionnaires. In this study, we investigate decoding performance of simultaneously induced cognitive and affective mental states with a filter-bank common spatial pattern (FBCSP) and linear discriminant analysis (LDA) approach [4,7]. In addition, we are interested how decoding performance changes when using subjectively rated labels instead of the experimentally induced labels.

Methods: We analyzed EEG data from five participants (2 female, 1 diverse; M = 23 years; SD = 1.02). Participants performed arithmetic tasks adding either 1-digit (low working memory load, LWML) or 2-digit numbers (high working memory load, HWML; cf., [8-9]; see [8] for a detailed description of the experimental task). Simultaneously, auditory emotional distractions of low (LV), neutral (NV), or high valence (HV) selected from the International Affective Digitized Sounds (IADS) database [10] were presented. Participants had to rate their emotional experience after each sound and perceived effort after adding a series of numbers of the same working memory load condition. EEG data was recorded using a mobile, dry-EEG with 20 electrodes (Cognionics Inc.; sampling rate: 500 Hz). EEG signals were de-trended, zero-padded, and re-referenced to mathematically linked mastoids [11]. We applied a zero-phase lag infinite impulse response (IIR) notch and finite impulse response (FIR) bandpass filter (cut-offs: 0.5 and 45 Hz). Signals were epoched in time intervals of 4 s starting at stimulus onset. Artefact removal comprised amplitude rejection (above 250 µV in any frontal electrodes) and an independent component analysis ([12-14] implemented in mne [15]). Finally, signals were downsampled (sampling rate: 250 Hz) and baseline corrected by subtracting the mean of a 200 ms time interval before stimulus onset. We extracted temporal-spatial discriminative features using a filter bank common spatial pattern (FBCSP) algorithm by bandpass-filtering the signal into multiple frequency bands (theta: 4 to 8 Hz, alpha: 8 to 12 Hz, low beta: 15 to 20 Hz, mid beta: 20 to 25 Hz). For the spatial CSP filters, we used 4 components with a regularized covariance estimation for each frequency-band [4,7]. We applied a principal component analysis (PCA) and selected the four most informative components via sequential forward feature selection (SFFS) nested for each subject and fold. To analyse the impact of subjective labels on decoding performance, we replaced the hypothetically assumed labels representing the conditions with subjectively rated labels by using a mean-split of the respective response scale and rerun the classification analysis (grand average congruency between experimental and subjective labels: 57.47 ± 17.06 % overall, 69.94 ± 17.48 % for emotional experience, and 82.48 ± 18.73 % for effort; see Table 1 for numbers of samples per class). Models were trained to classify conditions pairwise and in a four-class problem using a LDA classifier. As an empirical chance level, we trained a dummy classifier with randomly assigned labels. Classification performance was evaluated subject-wise within a stratified 3-fold cross-validation and balanced accuracy as metric. To gain a distribution of the average performance, we used a Monte Carlo Simulation (MCS) by retraining the ML pipeline 100 times.

Results and Discussion: Our results reveal that the FBCSP-LDA approach can predict the different combinations of affect and working memory load above chance when using the experimental condition as ground truth. Only for one participant above-chance classification could not be achieved when predicting working memory load under high valence distraction. Average balanced accuracy and 2.5th and 97.5th percentiles of the models and dummy classifiers are provided in Figure 1A and Table 1 (upper part) per subject and as grand averages (GA). Contrary to previous literature (c.f., [4]), we could successfully predict not only working memory load under different valence conditions but also valence under low and high working memory load. However, we could not predict conditions in a four-class classification. A potential explanation could be that emotional distractors induced further cognitive load, particularly under low working memory load, which dissolved distinctiveness between the working memory load conditions. The lower classification performance in one subject when predicting working memory load level under high valence distraction supports this suggestion. Regarding our second research question, we could not predict subjectively rated labels from neurophysiological signals (see Figure 1B and Table 1, lower part). The decrease in decoding performance might be explained by modulating effects, such as social desirability, cognitive dissonance for self-image maintenance, or limited ability to reliably estimate past experiences. To conclude, although average classification performance of our models was rather high when using the experimental condition as ground truth, we observed a high variability among folds. We suggest that further research with larger sample size is needed to: 1) gain more insights into possible explanations for the discrepancy between experimentally assumed and subjectively experienced states, 2) obtain suitable ground truths and calibration tasks for BCI training models that are valid to estimate mental states, and 3) further improve prediction pipelines and algorithms to ensure a robustly high accuracy to decode (interacting) mental states in enriched naturalistic environments.

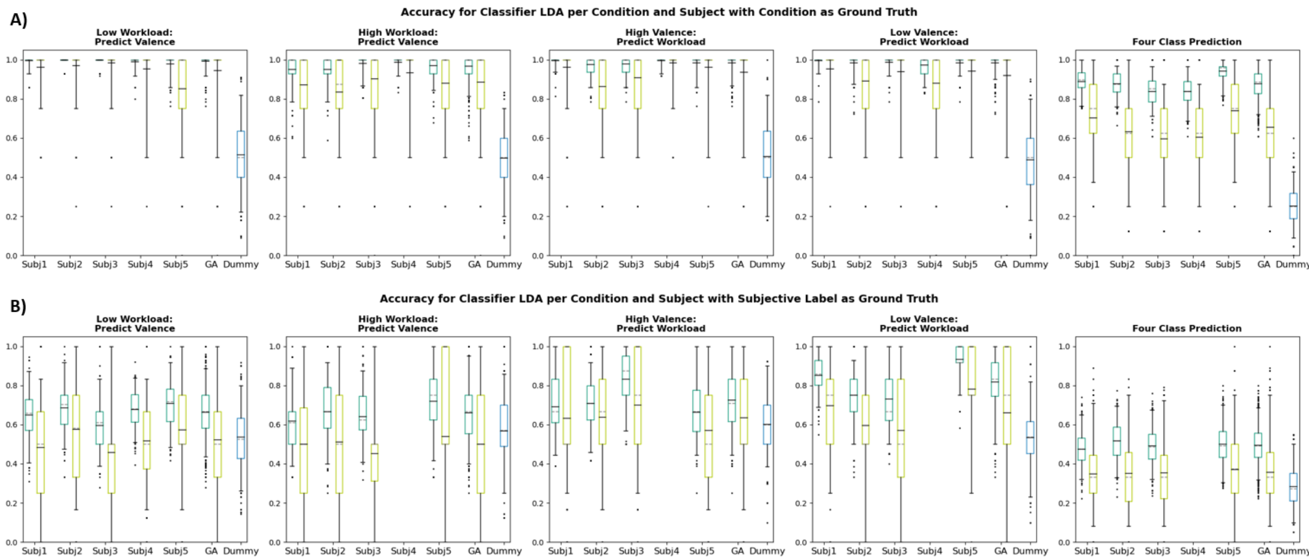

Figure 1. A) Decoding performance using FBCSP and a LDA classifier with the experimental condition as ground truth. B) Decoding performance with the subjective labels as ground truth. Distribution of balanced accuracy of the train (left box, green) and test set (right box, light green) averaged over 100 5-fold iterations and participants. Since Subj4 provided no ratings in the high working memory load range, only classification between the valence conditions with low working memory load was possible. Workload: here working memory load. Whiskers represent the 2.5 and 97.5th CI of the distribution. Solid line within the boxes: mean. Dashed line within the boxes: median. GA: Grand Average. Outer right box: grand average performance of the dummy classifier as empirical chance level (trained subject-wise on test set; blue).

Table 1. Decoding performance of the LDA and dummy classifiers per subject and classification problem as well as grand average over subjects for the a) experimental condition (upper part) and b) subjective label as ground truth (lower part). HV: High Valence. LV: Low Valence. HWML: High Working Memory Load. LWML: Low Working Memory Load. M: Mean. SD: Standard Deviation. Yellow: Above chance level prediction with no overlap of the 2.5 and 97.5th percentiles of the model and dummy classifier. Red: Above chance level prediction with partial overlap without including means of the model and dummy classifier.

Ground Truth Condition

Classification Problem	Participant	Subj1	Subj2	Subj3	Subj4	Subj5	Grand Average (GA)
Low Working memory load: Predict Valence HV – LV LWML	**Test performance**	0.962 [0.750; 1.000]	0.971 [0.750; 1.000]	0.985 [0.750; 1.000]	0.953 [0.500; 1.000]	0.852 [0.500; 1.000]	0.944 [0.500; 1.000]
	Dummy performance	0.457 [0.136; 0.786]	0.457 [0.136; 0.750]	0.452 [0.143; 0.818]	0.457 [0.150; 0.778]	0.440 [0.100; 0.800]	0.453 [0.136; 0.800]
	Samples (Class 1 – Class 2)	27 – 26	27 – 26	25 – 25	25 – 20	24 – 25	25.60 (±1.2) – 24.40 (±2.24)
High Working memory load: Predict Valence HV – LV HWML	**Test performance**	0.873 [0.500; 1.000]	0.836[0.500; 1.000]	0.903 [0.500; 1.000]	0.935 [0.500; 1.000]	0.881 [0.500; 1.000]	0.886 [0.500; 1.000]
	Dummy performance	0.457 [0.136; 0.786]	0.457 [0.136; 0.750]	0.452 [0.143; 0.818]	0.457 [0.150; 0.778]	0.440 [0.100; 0.800]	0.453 [0.136; 0.800]
	Samples (Class 1 – Class 2)	28 – 28	28 – 28	28 – 27	26 – 25	24 – 25	26.80 (±1.6) – 26.60 (±1.36)

High Valence: Predict Working memory load HV LWML – HWML	**Test performance**	0.963 [0.750; 1.000]	0.865 [0.369; 1.000]	0.909 [0.500; 1.000]	0.984 [0.750; 1.000]	0.963 [0.500; 1.000]	0.937 [0.500; 1.000]
	Dummy performance	0.457 [0.136; 0.786]	0.457 [0.136; 0.750]	0.452 [0.143; 0.818]	0.457 [0.150; 0.778]	0.440 [0.100; 0.800]	0.453 [0.136; 0.800]
	Samples (Class 1 – Class 2)	27 – 28	27 – 28	25 – 28	25 – 26	24 – 24	25.60 (±1.2) – 26.80 (±1.6)
Low Valence: Predict Working memory load LV LWML –HWML	**Test performance**	0.953 [0.500; 1.000]	0.891 [0.500; 1.000]	0.939 [0.500; 1.000]	0.881 [0.369; 1.000]	0.943 [0.500; 1.000]	0.921 [0.500; 1.000]
	Dummy performance	0.457 [0.136; 0.786]	0.457 [0.136; 0.750]	0.452 [0.143; 0.818]	0.457 [0.150; 0.778]	0.440 [0.100; 0.800]	0.453 [0.136; 0.800]
	Samples (Class 1 – Class 2)	26 – 28	26 – 28	25 – 27	20 – 25	25 – 25	24.40 (±2.24) – 26.60 (±1.36)
Four Class Prediction LV HWML – HV HWML – LV LWML – HV LWML	**Test performance**	0.704 [0.375; 1.000]	0.634 [0.250; 1.000]	0.595 [0.250; 0.875]	0.606 [0.250; 0.875]	0.740 [0.375; 1.000]	0.656 [0.250; 1.000]
	Dummy performance	0.457 [0.136; 0.786]	0.457 [0.136; 0.750]	0.143 [0.452; 0.818]	0.457 [0.150; 0.778]	0.440 [0.100; 0.800]	0.453 [0.136; 0.800]
	Samples (Class 1 – Class 2 – Class 3 – Class 4)	28 – 28 – 26 – 27	28 – 28 – 26 – 27	27 – 28 – 25 – 25	25 – 26 – 20 – 25	25 – 24 – 25 – 24	26.60 (±1.36) – 26.80 (±1.60) – 24.40 (±2.24) – 25.60 (±1.20)

Ground Truth Subjective Label

Classification Problem	**Participant**	**Subj1**	**Subj2**	**Subj3**	**Subj4**	**Subj5**	**Grand Average (GA)**
Low Working: Predict Valence HV – LV LWML	**Test performance**	0.484 [0.000; 0.921]	0.578 [0.167; 1.000]	0.460 [0.000; 1.000]	0.519 [0.145; 0.833]	0.574 [0.167; 1.000]	0.523 [0.000; 1.000]
	Dummy performance	0.503 [0.167; 0.800]	0.488 [0.182; 0.793]	0.542 [0.200; 0.818]	0.494 [0.316; 0.712]	0.496 [0.158; 0.833]	0.507 [0.158; 0.800]
	Samples (Class 1 – Class 2)	34 – 27	45 – 24	35 – 12	50 – 45	36 – 23	40.0 (±6.36) – 26.20 (±10.68)
High Working memory load: Predict Valence HV – LV HWML	**Test performance**	0.503 [0.000; 1.000]	0.512 [0.000; 1.000]	0.454 [0.000; 1.000]	–	0.539 [0.000; 1.000]	0.502 [0.000; 1.000]
	Dummy performance	0.503 [0.167; 0.800]	0.488 [0.182; 0.793]	0.542 [0.200; 0.818]	–	0.496 [0.158; 0.833]	0.507 [0.158; 0.800]
	Samples (Class 1 – Class 2)	12 – 34	17 – 22	14 – 39	0 – 0	12 – 23	11.0 (±5.8) – 23.6 (±13.46)
High Valence: Predict Working memory load HV LWML – HWML	**Test performance**	0.632 [0.206; 1.000]	0.640 [0.167; 1.000]	0.702 [0.250; 1.000]	–	0.570 [0.079; 1.000]	0.636 [0.167; 1.000]
	Dummy performance	0.503 [0.167; 0.800]	0.488 [0.182; 0.793]	0.542 [0.200; 0.818]	–	0.496 [0.158; 0.833]	0.507 [0.158; 0.800]
	Samples (Class 1 – Class 2)	34 – 12	45 – 17	35 – 14	50 – 0	36 – 12	40.0 (±6.36) – 11.0 (±5.8)
Low Valence: Predict Working	**Test performance**	0.697 [0.250; 1.000]	0.596 [0.000; 1.000]	0.572 [0.000; 1.000]	–	0.783 [0.250; 1.000]	0.662 [0.000; 1.000]

memory load LV LWML –HWML	**Dummy performance**	0.503 [0.167; 0.800]	0.488 [0.182; 0.793]	0.542 [0.200; 0.818]	–	0.496 [0.158; 0.833]	0.507 [0.158; 0.800]
	Samples (Class 1 – Class 2)	27 – 34	24 – 22	12 – 39	45 – 0	23 – 23	26.20 (±10.68) – 23.6 (±13.46)
Four Class Prediction LV HWML – HV HWML – LV LWML – HV LWML	**Test performance**	0.349 [0.083; 0.716]	0.351 [0.083; 0.750]	0.356 [0.083; 0.722]	–	0.372 [0.000; 0.750]	0.357 [0.083; 0.750]
	Dummy performance	0.503 [0.167; 0.800]	0.488 [0.182; 0.793]	0.542 [0.200; 0.818]	–	0.496 [0.158; 0.833]	0.507 [0.158; 0.800]
	Samples (Class 1 – Class 2)	34 – 12 – 27 – 34	22 – 17 – 24 – 45	39 – 14 – 12 – 35	0 – 0 – 45 – 50	23 – 12 –23 –36	23.60 (±13.46) – 11.0 (±5.8) – 26.20 (±10.68) – 40.0 (±6.36)

References:

[1] Fairclough, S. H. (2009). Fundamentals of physiological computing. *Interacting with Computers* 21, 133–145. doi: 10.1016/j.intcom.2008.10.011

[2] Picard, R. W. (2000). *Affective computing*. Cambridge, Massachusetts, London: The MIT Press.

[3] Parasuraman, R., and Rizzo, M. (2008). *Human Technology Interaction Series: Neuroergonomics*. Oxford: Oxford University Press.

[4] Appriou, A., Cichocki, A., and Lotte, F. (2020). Modern Machine-Learning Algorithms: For Classifying Cognitive and Affective States From Electroencephalography Signals. *IEEE Syst. Man Cybern. Mag.* 6, 29–38. doi: 10.1109/MSMC.2020.2968638

[5] Vukelić, M., Lingelbach, K., Pollmann, K., and Peissner, M. (2020). Oscillatory EEG Signatures of Affective Processes during Interaction with Adaptive Computer Systems. *Brain Sci* 11. doi: 10.3390/brainsci11010035

[6] Vukelić, M. (2021). Connecting Brain and Machine: The Mind Is the Next Frontier. In Friedrich, O., Wolkenstein, A., Bublitz, C., Jox, R. J., and Racine, E. (2021). *Clinical Neurotechnology meets Artificial Intelligence: Philosophical, Ethical, Legal and Social Implications*. Cham: Springer International Publishing; Imprint: Springer.

[7] Ang, K. K., Chin, Z. Y., Zhang, H., and Guan, C. (2008). "Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface," in *The 2008 IEEE International Joint Conference on Neural Networks: IJCNN 2008] ; Hong Kong, China, 1 - 8 June 2008 ; [part of 2008 IEEE World Congress on Computational Intelligence (WCCI 2008)* (Piscataway, NJ: IEEE Service Center), 2390–2397.

[8] Gado, S., Lingelbach, K., Bui, M., Rieger, J. W., and Vukelić, M. (2021). "Real-time feedback of subjective affect and working memory load based on neurophysiological activity," in *The International Conference, HCI International 2021, Washington DC, USA, July 24-29, 2021, Proceedings* (in-press). Springer.

[9] Lingelbach, K., Gado, S., Vukelić, M, and Rieger, J. W. (2021). "Investigating the Emotion-Cognition Interaction: Effects of Affective Distractors on Working Memory Load," in *The 3rd International Neuroergonomics Conference, Munich, Germany, September 11-16, 2021,* (submitted).

[10] Bradley, M. M., and Lang, P. J. (2007). The International Affective Digitized Sounds (IADS-2): Affective ratings of sounds and instruction manual. *University of Florida, Gainesville, FL, Tech. Rep. B-3*.

[11] Nunez, P. L., and Srinivasan, R. (2006). *Electric fields of the brain: The neurophysics of EEG*. Oxford: Oxford Univ. Press.

[12] Chaumon, M., Bishop, D. V. M., and Busch, N. A. (2015). A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *J Neurosci Methods* 250, 47–63. doi: 10.1016/j.jneumeth.2015.02.025.

[13] Hipp, J. F., and Siegel, M. (2013). Dissociating neuronal gamma-band activity from cranial and ocular muscle activity in EEG. *Front Hum Neurosci* 7, 338. doi: 10.3389/fnhum.2013.00338

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

[15] Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2014). MNE software for processing MEG and EEG data. *Neuroimage* 86, 446–460. doi: 10.1016/j.neuroimage.2013.10.027

Acknowledgements

This work was supported by grants from the Baden-Wuerttemberg Ministry for Economic Affairs, Labour and Housing (Project »KI-Fortschrittszentrum Lernende Systeme und Kognitive Robotik«).