Decoding subjective emotional arousal during a naturalistic VR experience from EEG using LSTMs

1st Simon M. Hofmann
Amsterdam Brain and Cognition
University of Amsterdam
Amsterdam, Netherlands
simon.hofmann@pm.me

2nd Felix Klotzsche
Berlin School of Mind and Brain
Humboldt Universität zu Berlin
Berlin, Germany
klotzsche@cbs.mpg.de

3rd Alberto Mariola
Sussex Neuroscience
University of Sussex
Sussex, UK
alberto.mariola@gmail.com

4th Vadim V. Nikulin
Neurology
MPI Hum. Cog. & Brain Sci.
Leipzig, Germany
nikulin@cbs.mpg.de

5th Arno Villringer
Neurology
MPI Hum. Cog. & Brain Sci.
Leipzig, Germany
villringer@cbs.mpg.de

6th Michael Gaebler
Neurology
MPI Hum. Cog. & Brain Sci.
Leipzig, Germany
gaebler@cbs.mpg.de

***Abstract*— Emotional arousal (EA) denotes a heightened state of activation that has both subjective and physiological aspects. The neurophysiology of subjective EA, among other mind-brain-body phenomena, can best be tested when subjects are stimulated in a natural fashion. Immersive virtual reality (VR) enables naturalistic experimental stimulation and thus promises to increase the ecological validity of research findings i.e., how well they generalize to real-life settings. In this study, 45 participants experienced virtual rollercoaster rides while their brain activity was recorded using electroencephalography (EEG). A Long Short-Term Memory (LSTM) recurrent neural network (RNN) was then trained on the alpha-frequency (8-12 Hz) component of the EEG signal (input) and the retrospectively acquired continuous reports of subjective EA (target). With the LSTM-based model, subjective EA could be predicted significantly above chance level. This demonstrates a novel EEG-based decoding approach for subjective states of experience in naturalistic research designs using VR.**

Keywords—subjective experience, neural decoding, emotional arousal, continuous time series, naturalistic research designs

I. INTRODUCTION

Emotional arousal (EA) is a “core affect” preparing an agent to respond to events in its environment [1]. With both subjective and physiological components, it is of central interest for the mind and brain sciences. The subjective and bodily aspects of EA are actively studied [2] – also under naturalistic conditions: McCall et al. [3] demonstrated that retrospective reports of EA correlate with peripheral physiological responses (here: heart rate, HR; skin conductance, SCR) during a naturalistic and immersive virtual reality (VR) experience. However, brain processes related to EA have only been tested in research paradigms that used relatively simplistic stimuli: For instance, an electroencephalography (EEG) study associated higher arousal, induced through pictures and music, with decreased alpha oscillatory power (8-12 Hz) over parietal brain regions [4]. To generalize such results to real-life settings (i.e., to increase their ecological validity), more complex and naturalistic research designs are required. To this aim, EEG – one of the most mobile neuroimaging techniques – can be combined with VR head-mounted displays (HMDs). Many classical EEG studies repeatedly present a stimulus in a trial-by-trail design. By averaging neural responses over trials, they extract event-related potentials. By design, this creates an artificial experience for participants. To avoid this, we aimed to provide a continuous and coherent natural experience to extract relevant neural and subjective features of EA.

In contrast to other deep learning models, Long Short-Term Memory (LSTM) recurrent neural networks (RNNs) [5] are quick “learners” and thus a promising analysis technique for the continuous data recorded under naturalistic stimulation with VR to overcome the limitations of trial-by-trial designs. Recently, LSTMs have been widely applied in the processing of complex time-sequential data, such as speech recognition [6] or video analysis [7]. However, despite their power to detect both short- and long-term dependencies in such time series, they have been rarely used for EEG data [8, 9].

The goal of our study was to generalize previous findings on the neurophysiology of EA in the alpha-frequency band to more ecologically valid settings. Subjects underwent virtual rollercoaster rides while their EEG was recorded. They then continuously rated their previously experienced levels of EA while viewing a recording of their rides (cf. [3]). LSTM-based models were then used for affective decoding, that is, to predict subjective EA from the EEG’s alpha-frequency components.

II. METHODS

A. Participants

45 healthy participants (22 men, mean age: 23±4, range: 20-32 years) were tested, of which 38 were analyzed (18 men). Data of 5 participants were lost, 1 participant stopped the experiment and 1 violated inclusion criteria. Subjects were right-handed, had normal or corrected-to-normal vision, and reported no psychiatric or neurological history.

B. Materials

- EEG (sampled at 500 Hz, hardware-based low-pass filter at 131 Hz) was recorded with 30 active Ag/AgCl electrodes attached according to the international 10-20 system (*actiCap* and *LiveAmp*, *Brain Products GmbH,* Germany). Two additional electrodes captured eye movements.
- HR and SCR were synchronously recorded (sampled at 500 Hz) with additional electrodes.
- VR setup: *HTC Vive* HMD (*HTC*, Taiwan), with headphones, attached on top of the EEG cap using

978-1-5386-9269-1/18/$31.00 ©2018 IEEE
DOI 10.1109/AIVR.2018.00026

cushions to avoid pressure artifacts. The two rollercoasters are commercially available [10].

C. Experimental Procedure

Participants had a 280-s VR experience of two rollercoasters (153 s, 97 s), including an intermediate 30-s break (all three were considered part of one continuous experience). In a first condition, participants were instructed to keep their head straight to avoid movement-related artifacts in the EEG data. In a second condition, they could move their head freely. In the subsequent rating phase, subjects saw a 2D recording of their experience on a virtual screen. While viewing the video, subjects recalled their EA and continuously reported it using a dial (Griffin PowerMate USB; sampling frequency: 50 Hz), with which they manipulated a vertical rating bar next to the video, ranging from low (0) to high (50) EA (cf. [3]).

D. Data Preprocessing

- The data (EEG, retrospective reports) were cropped by 5 s to avoid outliers related to the onset and offset of the virtual roller coaster rides. This resulted in time series of 270 s (HR and SCR data were not considered here).
- Subjective reports were downsampled to 1 Hz and re-scaled to the [-1,1] range. For the binary classification, low and high arousal (see next section) were defined as lower and upper tercile of the ratings, respectively. Entries on the tercile boundaries were semi-randomly assigned, keeping the bins equal in size (n = 90). The middle bin was removed, resulting in 180 samples per subject.
- EEG recordings were downsampled to 250 Hz. Preprocessing was standardized and automated with the *PREP pipeline* [11]. Independent components related to eye and head movements were removed using *MARA* [12].
- The EEG signal was decomposed with *spatio-spectral decomposition* (SSD) [13], which emphasizes the bandwidth of interest (here: alpha, 8-12 Hz) while attenuating adjacent frequencies. This also reduces noise through muscle activation, which normally occurs in frequencies above the alpha range (i.e., ~20-300 Hz) [14].
- We then used *Source Power Comodulation* (SPoC) [15], a supervised decomposition algorithm that maximizes the correlation between the target variable (here: ratings) and the time course of the power (here: in alpha) in time-sequential neural data. Training LSTMs (see next section) on supervised SPoC components aimed to set a performance benchmark proxy for affective decoding models trained on the purely alpha-informed, unsupervised SSD components.

E. LSTM-based Neural Decoding Model

An LSTM cell has the property to store and control relevant information of time series, as those of EEG. This feature was used to predict subjective states of EA from neural alpha-frequency components in two ways: i) a binary classification task (BCT), and ii) a continuous prediction task (ConT). The former aids the extraction of corresponding neural features by simplifying data into dichotomic targets. The latter was expected to be more demanding due to the more fine-grained temporal resolution. For each task, the best model hyperparameters (HPs) were found with a random search strategy [16] on a random subset of 10 subjects. The model was constrained to maximally two LSTM layers followed by maximally two fully connected layers (FC). The pyramidal design (Fig. 1) constrains successive layers to be equal or smaller in size (range: 10-100 nodes). Further HPs were the between-layer activation functions: rectified (ReLU) or exponential linear units (ELU) and different weight regularization methods (L1, L2) of various strengths (range: 0.0-1.44) that were added to the mean-squared error loss. The number (1-10) and selection of neural components as well as their transformation (width of the band-pass frequency filter, Hilbert alpha power extraction [17]) were also treated as HPs.

Finally, for each subject's dataset, separate models (SSD-BCT, SSD-ConT, SPoC-BCT, SPoC-ConT) were trained with *10-fold cross-validation* [18]. The LSTM was fed with mini-batches of size 9, where each sample corresponded to one second of the experience (sample size: $N_{ConT} = 270$, $N_{BCT} = 180$). Consequently, the model ran over 250 data points of neural components before it would output its prediction for one second of experience (see Fig. 1). Weights were optimized with *Adam* [19]. At this stage, we restricted the analysis to the data from the non-movement condition, since head motion-related artifacts in the data could corrupt the development of the model.

The LSTMs-based affective decoding model was implemented in the *Python 3.5.1* package *Tensorflow 1.4.1* (*Google Inc.*, USA). All scripts are available online [20].

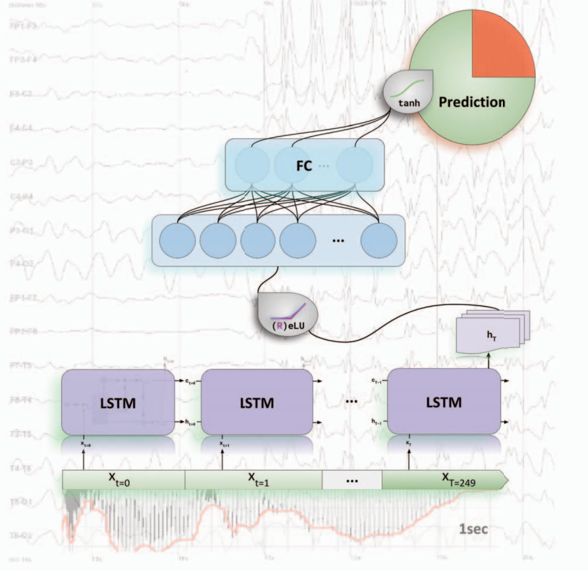

Fig. 1 Model architecture: The LSTMs (1-2 layers) were fed with 1-s slices of EEG recordings. The last hidden state (h_T) was channeled to the fully connected layer(s) (FC), which output the prediction through a tangens hyperbolicus (tanh).

III. Results

For both approaches (BCT, ConT), the performances of SSD- and SPoC-trained models were compared.

A. *Binary classification (BCT)*

- **SSD-trained models:** the mean accuracy to predict subjective ratings on the validation sets over all subjects was 0.634 (range: 0.514-0.816, sd = 0.068), which was significantly above chance level (perm$_{3000}$ $p < 0.001$). Fig. 2 shows one subject's prediction model.

- **SPoC-trained models:** the mean accuracy was 0.623 (range: 0.499-0.879, sd = 0.077), which was also significantly above chance level (perm$_{3000}$ $p < 0.001$).

- LSTMs trained on SSD neural components did not differ significantly from LSTMs trained on SPoC components (perm$_{3000}$ $p = 0.554$); hence, the benchmark proxy was reached. The accuracies also matched the results on the same dataset reported in [21], which used *common spatial patterns* (CSP) [22], a neural decoding algorithm primarily used for EEG-based brain-computer interfaces.

B. *Continuous prediction (ConT)*

- **SSD-trained models:** the mean prediction accuracy on the validation sets over all subjects was 0.757 (range: 0.677-0.825, sd = 0.036), which was significantly above the *average-line accuracy* (range: 0.591-0.820, sd = 0.051), that is, the accuracy that the model would achieve when it only outputs the average rating per subject (perm$_{3000}$ $p < 0.001$). Thus, the models could detect features in the neural data to decode subjective EA.

- **SPoC-trained models:** the mean accuracy was 0.754 (range: 0.679-0.826, sd = 0.035), which was also significantly above the average-line accuracy (perm$_{3000}$ $p < 0.001$).

- LSTMs trained on SSD neural components did not differ significantly from LSTMs trained on SPoC components (perm$_{3000}$ $p = 0.735$); hence, the benchmark proxy was reached.

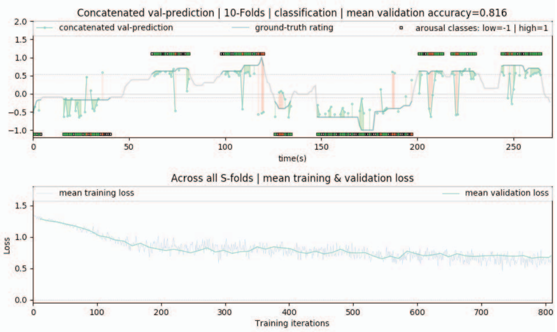

Fig. 2 Prediction of subjective rating over the 270-s VR experience (two rollercoasters and the intermediate break), and learning progress of the best SSD dataset in the binary classification (BCT) (Subject 23, mean validation accuracy = 0.816): Concatenated prediction on the validation set over 10 folds (*Top*). Training progress over 810 iterations (*Bottom*).

Across prediction approaches, different sets of HPs showed similar results. There was no clear trend neither for input transformations of neural data (choice of bandwidth filter, Hilbert power extraction) nor for between-layer activation functions nor for weight regularization methods being clearly beneficial for the final prediction accuracy on the validation set. However, more narrow networks with maximally three neural components tended to have higher accuracy rates. For both SSD- and SPoC-trained LSTMs, the best prediction outcomes were achieved on high-rank components. For SSD, the rank describes the strength of alpha information in the component. For SPoC, the rank describes the magnitude of comodulation between the target (here: rating) and the neural component.

In summary, all combinations of prediction tasks (SSD-BCT, SSD-ConT, SPoC-BCT, SPoC-ConT) showed accuracies significantly above chance (BCT) or the average-line accuracy level (ConT).

IV. Discussion

In an immersive VR-EEG study with virtual rollercoasters, we induced EA in subjects to decode their (retrospectively rated) emotional experience from neural responses. We found that LSTM-based affective decoding models can extract features from neural input components that reflect the subjective experience of EA. LSTMs thus provide a suitable analytical approach for complex time series (e.g., brain activation measured using EEG) from naturalistic stimulation (e.g., using immersive VR). Such research designs are needed to increase the ecological validity of findings in the mind and brain sciences.

We focused on neural features in the alpha-frequency range that were previously found to correlate negatively with arousal in less naturalistic experiments [4]. Although we here did not aim at a general understanding of neural oscillations during arousing experiences, this could be investigated by applying our decoding model (code at [20]) to other frequency bands. Since the LSTM was trained on neural features extracted with SSD and SPoC, the topological distribution of oscillations could be analyzed by reprojecting the corresponding filter matrices [23], as it was done with the CSP approach [21]. Our implementation would also allow training the LSTM model with other physiological modalities such as HR and SCR (cf. [3]). However, different signals (e.g., HR and neural recordings) usually have different temporal properties, which might require adapting the HPs and lead to computationally expensive HP searches. Systematically analyzing the performance as a function of model architecture could be informative about how fruitful an exploration of the search space would be. Our results suggest that this variance is not substantial after an initial *broad* search that promotes a pre-selection of HP sets.

Although our model could decode subjective states from neural recordings, the accuracy was not as high as in other EEG-based paradigms, such as lateralized motor-imagery classification [24]. This may be due to i) the rapidly changing events of the virtual rides, ii) the participants' retrieval of the corresponding emotional states from memory during the rating phase, iii) the variability of subjective reports in general [25], and/or iv) the one-trial study design and its relatively short time series.

The modality of the model input (SSD, SPoC, various transformations such as the Hilbert transformation) did not significantly affect the model performance. This corroborates findings that deep learning models are effective function approximators [26]. Similarly, convolutional neural networks trained on EEG data approximate particular processes of features extractions for neural data [27].

In future experiments, it should be tested whether our affective decoding model can also extract neural features *across* subjects. So far, we trained models on datasets of single subjects. Even though these datasets are relatively short, the LSTM was able to learn quickly and converged early in the training process (Fig. 2). This counters the often-stated drawback of deep learning models to require huge datasets, which lead to extensive training periods. Here, the benefit of LSTMs over other deep learning models is their fast gradient-flow through their memory cells during the weight-update [5]. Since this type of algorithm remains difficult to interpret, we only fed the model with signal features of interest (in this case: alpha-frequency components). In future work, this approach could be compared to an end-to-end learning (i.e., training the model on raw data). However, end-to-end learning could artificially increase or decrease the model performance due to systematic artifacts related to non-neural (e.g., muscular) activation. While profiting from the abilities of LSTMs and deep learning models in general, it is essential to simultaneously develop new methods for a better interpretation of their processes (cf. [28]), which would support training models on raw data. Ultimately, as the state of an agent is also a function of its environment, the integration of features of the (VR) stimulus itself into the model could be a crucial step towards a better understanding of the multifaceted phenomenon of arousal.

V. Conclusion

The subjective experience of EA has previously been linked to peripheral physiological states in an immersive VR experiment [3]. In a non-VR study with more simple stimuli, oscillations in the alpha frequency range (8-12 Hz) were found to decrease during higher arousal [4]. We combined both approaches, showing that subjective EA can be successfully predicted from alpha oscillations in a setup that measures EEG during an immersive VR experience.

We conclude that LSTM RNNs can be used to decode subjective experience from neural information in naturalistic research designs. We hope that such stimulation combining immersive VR and neuroimaging may provide a tool to increase the ecological validity of neuroscientific experiments.

Acknowledgment

Thanks to Wojciech Samek, Klaus Robert-Müller, Frederik Harder, Kristof Schütt, and Leila Arras, Cade McCall, Alireza Tarikhi, Mert Akbal, Firat Sansal.

References

[1] J. A. Russell and L. Feldman Barrett, “Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant,” J. Pers. and Soc. Psychol., vol. 76, no. 5, 1999.

[2] I. B. Mauss, R. W. Levenson, L. McCarter, F. H. Wilhelm, and J. J. Gross, “The tie that binds? Coherence among emotion experience, behavior, and physiology,“ Emotion, vol. 5, no. 2, 2005.

[3] C. McCall, L. K. Hildebrandt, B. Bornemann, and T. Singer, “Physiophenomenology in retrospect: Memory reliably reflects physiological arousal during a prior threatening experience,” Consciousness and Cognition, vol. 38, 2015, pp. 60–70.

[4] C. Di Bernardi Luft and J. Bhattacharya, “Aroused with heart: Modulation of heartbeat evoked potential by arousal induction and its oscillatory correlates,” Scientific Reports, vol. 5, no. 15717, 2015.

[5] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Computation, vol. 9, no. 8, 1997, pp. 1735–1780.

[6] A. Graves, N. Jaitly, and A.-R. Mohamed, “Hybrid speech recognition with deep bidirectional LSTM,” in Automatic Speech Recognition and Understanding, 2013, pp. 273– 278.

[7] J. Donahue, L. Anne Hendricks, S. Guadarrama et al. “Long-term recurrent convolutional networks for visual recognition and description,” IEEE CVPR, 2015, pp. 2625-2634.

[8] P. Bashivan, G. M. Bidelman, and M. Yeasin, “Spectrotemporal dynamics of the EEG during working memory encoding and maintenance predicts individual behavioral capacity,” Eur. J. Neurosci., vol. 40*,* 2014, pp. 3774–3784.

[9] R. G. Hefron, B. J. Borghetti, J. C. Christensen, and C. M. Schubert, “Deep long short-term memory structures model temporal dependencies improving cognitive workload estimation,” Pattern Recognition Letters, vol. 94, 2017, pp. 96–104.

[10] Funny Twins Games, “Russian VR Coasters,” Russia, 2016.

[11] N. Bigdely-Shamlo, T. Mullen, C. Kothe, K.-M. Su, and K. A. Robbins, “The PREP pipeline: standardized preprocessing for large-scale EEG analysis,” Frontiers in Neuroinformatics, vol. 9, June 2015.

[12] I. Winkler, S. Brandl, F. Horn et al. “Robust artifactual independent component classification for BCI practitioners,” J. Neural Eng., vol. 11, no. 3, 2014, pp. 1–11.

[13] V. V. Nikulin, G. Nolte, and G. Curi, ”A novel method for reliable and fast extraction of neuronal EEG/MEG oscillations on the basis of spatio-spectral decomposition,” NeuroImage, vol. 55, no. 4, 2011.

[14] S. D. Muthukumaraswamy, “High-frequency brain activity and muscle artifacts in MEG/EEG: A review and recommendations,” Frontiers in Human Neuroscience, vol. 7, no. 138, 2013.

[15] S. Dähne, F. C. Meinecke, S. Haufe et al. “SPoC: A novel framework for relating the amplitude of neuronal oscillations to behaviorally relevant parameters,” NeuroImage, vol. 86, 2014, pp. 111–122.

[16] J. Bergstra and Y. Bengio, “Random search for hyper-parameter optimization,” J. Mach. Learn. Res., vol. 13, 2012, pp. 281–305.

[17] M. X. Cohen, “Analyzing neural time series data: Theory and practice,” Cambridge, Massachusetts: MIT press, 2014.

[18] C. M. Bishop, “Pattern Recognition and Machine Learning,” 4th ed., New York: Springer, 2006.

[19] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” ICLR, 2015.

[20] S. M. Hofmann, “NeVRo,” GitHub Repository, 2017, [Online] Available: https://github.com/SHEscher/NeVRo.

[21] F. Klotzsche, A. Mariola, S. M. Hofmann et al. “Using EEG to decode subjective levels of emotional arousal during an immersive VR roller coaster ride,” IEEE VR, 2018.

[22] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Müller, “Optimizing spatial filters for robust EEG single-trial analysis,” IEEE Signal processing magazine, vol. 25, no. 1, 2008, pp. 41-56.

[23] S. Haufe, S. Dähne, and V. V. Nikulin, “Dimensionality reduction for the analysis of brain oscillations,” NeuroImage, 2014.

[24] P. Herman, G. Prasad, T. M. McGinnity, and D. Coyle, “Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification,” IEEE Trans. Neural Syst. Rehabil. Eng., vol. 16, 2008, pp. 317– 326.

[25] J. Blascovich, “Individual differences in physiological arousal and perception of arousal,” Pers. Soc. Psychol. Bull., vol. 16, no. 4, 1990.

[26] S. Liang and R. Srikant, “Why deep neural networks for function approximation?,” ICLR, 2017.

[27] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer et al. “Deep learning with convolutional neural networks for EEG decoding and visualization,” Human Brain Mapping, vol. 38, 2017, pp. 5391–5420.

[28] L. Arras, G. Montavon, K. R. Müller, W. Samek, “Explaining recurrent neural network predictions in sentiment analysis,” arXiv: 1706.07206v2 [cs.CL], August 2017.