2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Electroencephalography in Evaluating Mental Workload of Gaming

Ville Ahonen[1], Marko Leino[2] and Tarmo Lipping[3]

Abstract— **The feasibility of electroencephalography (EEG) analysis in evaluating mental workload of gaming was studied by carrying out a proof-of-concept type experiment on a set of EEG recordings, with a bespoke tool developed for the purpose. The EEG recordings (20 recordings in total) that were used in the experiment had been acquired by groups of students and staff of Tampere University during n-back gaming sessions, as part of course projects. The ratio of theta and alpha power, calculated over the EEG signal segments that were time-locked to game events, was selected as EEG metrics for mental load evaluation. Also, Phase Locking Value (PLV) was calculated for all pairs of EEG channels to assess the change in phase synchronization with the increasing difficulty level of the game. Wilcoxon rank-sum test was used to compare the metrics between the levels of the game (from 1-back to 4-back). The rank-sum test results revealed that the theta-alpha power ratio calculated from the frontal derivations Fp1 and Fp2 performed as a confident indicator for the evaluation and comparison of mental load. Also, phase locking between EEG derivations was found to become stronger with the increasing difficulty level of the game, especially in the case of channel pairs where the electrodes were located at opposite hemispheres.**

I. INTRODUCTION

Mental workload can be considered as an objective task demand imposed on a person's cognitive resources. Cognitive load theory (CLT) provides a more profound theoretical framework that is based on the cognitive architecture comprised of a working memory and a long-term memory, and that specifies mental workload according to its origins (intrinsic, extraneous or germane workload). The measures for mental workload can be divided into the categories of subjective measures, performance measures and psychophysiological measures. Electroencephalography (EEG) represents an indicator for the last category. Also, EEG provides convenient means for monitoring and evaluating mental load, without causing any significant interference to the subject under study [1]–[4]. EEG metrics for mental load assessment can be calculated from power spectrum, event related potential [5] or brain connectivity measurements. Gaming quite perfectly complements the EEG in the formation of a framework for studying the mental load as, in general, load imposing conditions in a game can be easily adjusted, and basic setups for such study environments are rather simple [6], [7].

[1]Ville Ahonen was with the Faculty of Information Technology and Communication Sciences until 2020, Tampere University, Pori, Finland

[2]Marko Leino is with the Faculty of Information Technology and Communication Sciences, Tampere University, Pori, Finland marko.leino@tuni.fi

[3]Tarmo Lipping is with the Faculty of Information Technology and Communication Sciences, Tampere University, Pori, Finland tarmo.lipping@tuni.fi

While variables such as heart rate, blood pressure or skin impedance assess the short-term immediate reaction to the increase of task difficulty [8], direct monitoring of the function of central nervous system using EEG, for example, is required to evaluate long-term mental workload and detect changes leading to burnout or depression.

II. MATERIALS AND METHODS

In order to determine mental workload from EEG responses, the n-back memory game was deemed as an appropriate test setting [4], [5], [9]. A brief description of the game is provided in Section II-A, and the details of the measurement setup and procedure are described in Section II-B.

A. N-back game

The n-back memory game was introduced by Kirchner [10] and it has been a widely used tool since then in numerous cognitive performance related studies, especially in those related to working memory performance and mental load.

In the n-back memory game a person is presented sequential stimuli that are perceptually identifiable, e.g. visual stimuli consisting of the set of alphanumerical characters or auditory stimuli consisting of the set of auditory probes. The person playing the game is requested to respond, e.g. by pressing a button, when the current stimulus being presented was presented also n items back. The matching stimulus is called the target whereas a non-matching stimulus is called a non-target. The outcome that reflects the task performance, like the number of correct and incorrect responses, as well as the response delay, can be then used by an experimenter to further analyze the cognitive performance. Game difficulty is controlled with the value of n, i.e. the higher the value the more difficult the game becomes as it imposes higher load on the working memory. When the value of n is zero, the stimulus used as a target is predefined.

B. Setup

The EEG measurements and stimuli related data used in this experiment were acquired by groups of Tampere University students and staff, as a part of course projects during the years 2019 and 2020. The test setup for course projects comprised the following components: a laptop, n-back game and EEG recording software running on the laptop, and an EEG electrode cap with a measurement unit connected wirelessly to the laptop.

The version of n-back game used in this experiment employed simple visual stimuli as single digits (0-9) displayed

TABLE I

6-CHANNEL AND 19-CHANNEL SETUPS

Setup	Connected EEG channels
6-channel	Fp1, Fp2, C3, C4, O1, O2
19-channel	Fp1, Fp2, F7, F3, Fz, F4, F8, T7, C3, Cz, C4, T8, P7, P3, Pz, P4, P8, O1, O2

on the computer screen in 3 second intervals. The subjects were supposed to click the mouse button when a displayed number was the target. Gaming sessions were always started with 1-back, proceeding up to 4-back, with a short break between the levels. The duration for each level was a few minutes. All game events during the sessions, i.e. n-back game level changes, displayed numbers and mouse clicks, were stored into a log file.

The EEG during the n-back gaming sessions, was measured and recorded with the Enobio® (Neuroelectrics, Spain) EEG system (see [11] for more information). The version of the measurement device that was used in the course projects carried out in 2019 supported up to 8 channels, and the version used in the projects in 2020 supported up to 20 channels.

In total, the projects covered 20 n-back gaming sessions, for which the EEG measurements and game events related data were recorded. A setup with 6 connected EEG channels was used in 18 of these sessions, and a higher density setup with 19 connected EEG channels in two of the sessions. The EEG signal sampling rate was 500 Hz for all recordings, and in all cases the ECG was recorded simultaneously with the EEG. The connected EEG channels in both the 6-channel and 19-channel setups are shown in Table I. The channels in the 6-channel setup are a subset of the channels in the 19-channel setup.

C. EEG processing

A MATLAB (version R2019a) [12] based tool was created for processing EEG recordings and for calculating EEG metrics — for the mental load evaluation — from the processed recordings. The source code for the tool can be found in the GitHub repository [13]. The tool was divided into two main level logical entities: the validator covering the initial processing of EEG recordings and the calculator covering the subsequent EEG metrics calculation from the processed EEG recordings. All EEG channels in all EEG recordings were filtered, cleaned and validated, in this order, by the respective algorithms that are described in the following.

Built-in MATLAB functions *firpmord* and *firpm* are applied in the implemented filtering algorithm (PMFilter, see [13]) to design a bandpass filter. The bandpass filter was deployed with the lower cut-off frequency set at 1 Hz and the upper cut-off frequency at 44 Hz. The artefact cleaning algorithm (AlgWaveletCleaner, see [13]) employs discrete wavelet transform (DWT) based multiresolution analysis (MRA), by the application of MATLAB functions *wfilters* and *modwt*. The cleaning algorithm was deployed with the low-pass and high-pass decomposition and reconstruction filters associated with the MATLAB pre-defined wavelet "db2" from the Daubechies wavelet family. With this setting the AlgWaveletCleaner algorithm executes MRA decomposition for the EEG signals, and thresholding for the decomposed detail components. The thresholding is carried out in the manner that samples of the detail components are zeroed, if they have an amplitude higher than 2.5 times the standard deviation calculated over the sample window of duration of 10 seconds and which is moved in 5 second steps. After the thresholding, the inverse DWT, using the MATLAB function *imodwt*, is performed to reconstruct the (cleaned) EEG signal.

After filtering and cleaning, the validation algorithm (AlgChannelValidator, see [13]) loops through the channel specific EEG signal segments, or epochs, and compares the average power of each epoch to the average power of the whole channel. If the average power of an epoch is more than two standard deviations away from the average channel power, the epoch is marked as invalid and will be omitted in the EEG metrics calculation. The epoch length was set to one second with 50% overlap.

D. Mental workload metrics

Two kinds of EEG metrics were calculated: the ratio of signal power in theta frequency band to that of alpha frequency band (Theta-Alpha Power Ratio, TAPR) and the Phase Locking Value (PLV) between channel pairs. The TAPR measure follows the Cognitive Load Index (CLI) method introduced by Holm et al. [14] with the difference that the ratio is calculated for each channel separately while in CLI two channels, Fz and Pz are involved. The metrics are calculated over signal segments that are time locked to game events, starting 100 ms before a number was displayed with the duration of 1000 ms. The power spectral density (PSD) estimate was calculated using the Welch's method (MATLAB function *pwelch*). Hamming window of length 128 samples with the overlap of 64 samples was used.

The phase locking value is a measure of synchronicity between the instantaneous phases of two time series. It was introduced for brain signal analysis by Lachaux et al. [15]. Prior to calculating the PLV, the EEG signal is filtered using a narrow-band band-pass filter so that PLV of different EEG rhythms can be assessed. The metric was calculated in the following steps (see [16]):

1) Signal segments of EEG channels i and j were band-pass filtered with passband cutoff frequencies f_1 and f_2 to obtain $x_i^{f_1,f_2}$ and $x_j^{f_1,f_2}$
2) The instantaneous phases of the narrow-band signals were obtained using the Hilbert transform: $\angle x_i^{f_1,f_2} = \varphi\left\{\mathbf{H}\left\{x_i^{f_1,f_2}\right\}\right\}$ and $\angle x_j^{f_1,f_2} = \varphi\left\{\mathbf{H}\left\{x_j^{f_1,f_2}\right\}\right\}$
3) PLV was calculated as:

$$\mathrm{PLV}_{x_i x_j}^{f_1,f_2} = \left|\mathbf{Avg}\left\{\exp\left\{\sqrt{-1}\left(\angle x_i^{f_1,f_2} - \angle x_j^{f_1,f_2}\right)\right\}\right\}\right|,$$

where $\mathbf{Avg}\{\cdot\}$ denotes averaging over the time samples of the signal segment.

In our study we calculated PLV for frequency bands $2-4$ Hz, $4-6$ Hz, $6-8$ Hz, $8-10$ Hz, $10-12$ Hz, $12-16$ Hz and $16-20$ Hz. The most significant results were obtained for the frequency band $2-4$ Hz.

III. RESULTS

TAPR values, related to n-back game events, were calculated per n-back game level for each channel in each recording. Outliers were removed by employing the median absolute deviation (MAD) method [17]. TAPR values more than three MAD values away from the median were considered as outliers. The channel specific TAPR values for the same n-back game level were combined from all recordings and are shown in Figure 1. The left-sided Wilcoxon rank-sum test was then separately performed for each pair of the n-back levels of the combined TAPR values. The p-values corresponding to the obtained rank-sum statistics are listed in Table II.

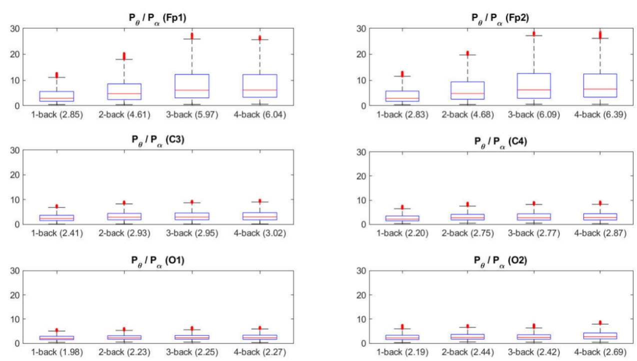

Fig. 1. Theta-alpha power ratio (TAPR) of all channels of combined recordings, for n=1,2,3,4. The median of TAPR for each n-back level is given in parenthesis.

It can be seen from Figure 1 that the difference between the medians of TAPR values for different n-back levels are highest in the Fp1 and Fp2 channels, whereas the medians in all the other channels lie on an almost flat line. This better TAPR performance for the Fp1 and Fp2 channels is evident by statistically significant p-values, at the significance level 0.001, for TAPR comparison between each n-back level, except between the 3-back and 4-back (Table II).

In Figure 2 the results for the PLV analysis are presented. The difference in the PLV metric between 1-back and 3-back levels of the game are presented for frequency band $2-4$ Hz. In the figure, channel pairs for which the p-value obtained using the Wilcoxon's test between the two game levels was below the threshold (indicated above each panel) are connected. On the left panel, the combined results for all the recordings are presented at two significance levels, $p \leq 0.01$ and $0 \leq 0.0001$, for 6 EEG channels while the right panel presents similar results for the two 19-channel recordings. For the clarity of representation, on the right panel tighter significance level thresholds $p \leq 0.0001$ and $p \leq 10^{-10}$ are used.

TABLE II

P-VALUES FOR TAPR COMPARISON BETWEEN DIFFERENT N-BACK LEVELS (LEFT-SIDED WILCOXON RANK-SUM TEST)

Ch	1vs2	1vs3	1vs4	2vs3	2vs4	3vs4
Fp1	<0.001	<0.001	<0.001	<0.001	<0.001	0.177
Fp2	<0.001	<0.001	<0.001	<0.001	<0.001	0.055
C3	<0.001	<0.001	<0.001	0.203	0.217	0.520
C4	<0.001	<0.001	<0.001	0.451	0.825	0.884
O1	<0.001	<0.001	<0.001	0.108	0.404	0.839
O2	0.048	0.001	<0.001	0.070	<0.001	<0.001

IV. DISCUSSION AND CONCLUSIONS

To shortly summarize the results of the analysis, the TAPR measure for the frontal EEG channels Fp1 and Fp2 seems to provide a plausible EEG indicator for mental load evaluation and comparison. This is aligned with the earlier discussed studies (see [14]) that evinced the theta synchronization and alpha desynchronization with increasing mental load. The difference is most significant and consistent between game difficulty levels 1 and 3 or 1 and 4. The reason why the TAPR value do not change significantly between game levels 3 and 4 might be that at level 4 the recalling the occurrence of the target stimulus becomes too demanding and the subjects tend to loose focus in the game.

To our knowledge, the results for the Phase Locking Value between EEG derivations during a n-back memory game have not been published previously. In [18] it has been shown that in the case of Action Real-time Strategy Gaming (ARSG) connections between the temporal and the central area of the brain as measured using the PLV metric were strengthened in comparison to the resting condition. Our results confirm these finding in general. Based on Figure 2 a couple of observations can be made. Firstly, most of the highly significant increases of the PLV occur between channel pairs of opposite hemispheres. Secondly, two separate networks seem to be involved in the coupling: one between the temporal and frontal areas with less significance and another involving central, parietal and occipital regions with high significance.

Our results may well correlate with the phenomenon shown in [19] that the power of the delta rhythm increases with increasing cognitive load. The increasing delta power is suggested to indicate the inhibition of the activities not related with the cognitive task. The author also suggests that the delta rhythm would modulate the activity of the networks that should be inactive during the task. Our finding of strengthening synchronisation between the delta rhythm of certain brain areas is complementary to the conclusion of [19]. The modulation of other activities by the delta rhythm need still to be shown.

However, in an attempt to obtain more consistent and less divergent TAPR results, more recordings should be performed in a better controlled environment. Also, other methods for artefact removal should be studied and experimented, including a visual assessment by an expert.

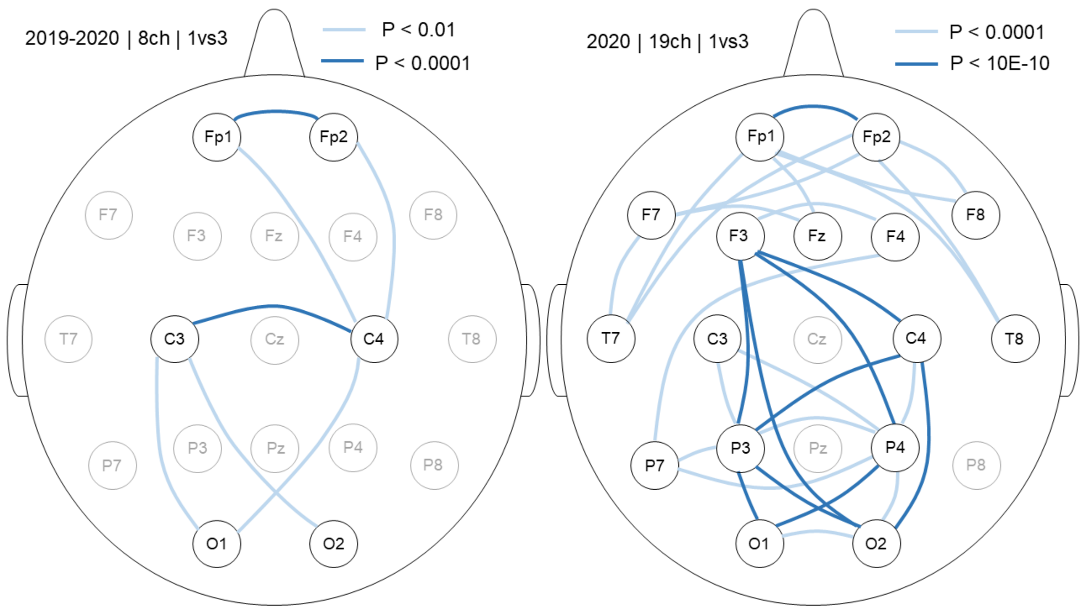

Fig. 2. Significance of the difference between the Phase Locking Value, calculated over all pairs of EEG channels, between 1-back and 3-back levels of the game.

REFERENCES

[1] G. N. Dimitrakopoulos, I. Kakkos, Z. Dai, J. Lim, J. J. Desouza, A. Bezerianos, and Y. Sun, "Task-Independent Mental Workload Classification Based Upon Common Multiband EEG Cortical Connectivity," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 1940–1949, 2017.

[2] W. K. So, S. W. Wong, J. N. Mak, and R. H. Chan, "An evaluation of mental workload with frontal EEG," *PLoS ONE*, vol. 12, no. 4, pp. 1–17, 2017. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0174949

[3] J. W. Ahn, Y. Ku, and H. C. Kim, "A novel wearable EEG and ECG recording system for stress assessment," *Sensors (Switzerland)*, vol. 19, no. 9, pp. 1–14, 2019.

[4] H. Aghajani and A. Omurtag, "Assessment of mental workload by EEG+FNIRS," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2016-Octob, pp. 3773–3776, 2016.

[5] A. M. Brouwer, M. A. Hogervorst, J. B. Van Erp, T. Heffelaar, P. H. Zimmerman, and R. Oostenveld, "Estimating workload using EEG spectral power and ERPs in the n-back task," *Journal of Neural Engineering*, vol. 9, no. 4, pp. 1–14, 2012.

[6] H. Aliyari, H. Sahraei, M. R. Daliri, B. Minaei-Bidgoli, M. Kazemi, H. Agaei, M. Sahraei, S. M. A. S. Hosseini, M. M. Hadipour, M. Mohammadi, and Z. Dehghanimohammadabadi, "Research Paper: The beneficial or harmful effects of computer game stress on cognitive functions of players," *Basic and Clinical Neuroscience*, vol. 9, no. 3, pp. 177–186, 2018.

[7] D. J. Krusienski, C. Tremmel, C. Herff, T. Sato, K. Rechowicz, and Y. Yamani, "Estimating Cognitive Workload in an Interactive Virtual Reality Environment Using EEG," *Frontiers in Human Neuroscience*, vol. 13, no. 4, p. 1, 2019.

[8] A. M. Porter and P. Goolkasian, "Video games and stress: How stress appraisals and game content affect cardiovascular and emotion outcomes," *Frontiers in Psychology*, vol. 10: 967, pp. 1–13, May 2019.

[9] M. S. Caywood, D. M. Roberts, J. B. Colombe, H. S. Greenwald, and M. Z. Weiland, "Gaussian process regression for predictive but interpretable machine learning models: An example of predicting mental workload across tasks," *Frontiers in Human Neuroscience*, vol. 10: 647, pp. 1–19, January 2017.

[10] W. K. Kirchner, "Age differences in short-term retention of rapidly changing information," *Journal of experimental psychology; J Exp Psychol*, vol. 55, no. 4, pp. 352–358, 1958.

[11] Enobio, "Wireless medical grade EEG systems for high precision EEG monitoring," 2020. [Online]. Available: https://www.neuroelectrics.com/solutions/enobio

[12] MathWorks, "MATLAB," 2019. [Online]. Available: https://se.mathworks.com/products/matlab.html

[13] V. Ahonen, "EEG-validator-calculator source code," 2020. [Online]. Available: https://github.com/vahonen/eeg-validator-calculator

[14] A. Holm, K. Lukander, J. Korpela, M. Sallinen, and K. M. Müller, "Estimating brain load from the EEG," *TheScientificWorldJournal*, vol. 9, pp. 639–651, jul 2009. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/19618092/

[15] J. P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela, "Measuring phase synchrony in brain signals," *Human Brain Mapping*, vol. 8, no. 4, pp. 194–208, 1999.

[16] T. Lipping, N. Erkintalo, M. Särkelä, R. S. K. Takala, A. Katila, J. Frantzén, J. P. Posti, M. Müller, and O. Tenovuo, "Connectivity analysis of full montage eeg in traumatic brain injury patients in the icu," in *EMBEC & NBC 2017*, H. Eskola, O. Väisänen, J. Viik, and J. Hyttinen, Eds. Singapore: Springer Singapore, 2018, pp. 97–100.

[17] C. Leys, C. Ley, O. Klein, P. Bernard, and L. Licata, "Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median," *Journal of Experimental Social Psychology*, vol. 49, no. 4, pp. 764–766, jul 2013. [Online]. Available: /record/2013-13220-001

[18] D. Gong, Y. Li, Y. Yan, Y. Yao, Y. Gao, T. Liu, W. Ma, and D. Yao, "The high-working load states induced by action real-time strategy gaming: An eeg power spectrum and network study," *Neuropsychologia*, vol. 131, pp. 42–52, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0028393219301113

[19] T. Harmony, "The functional significance of delta oscillations in cognitive processing," *Frontiers in integrative neuroscience*, vol. 7, p. 83, 2013.