Mental Flow Estimation Through Wearable EEG

Manuel Cherep[1], Mikolaj Kegler[2], Jean-Philippe Thiran[3] and Pablo Mainar[1]

***Abstract*— Flow is a mental state experienced during holistic involvement in a certain task, and it is a factor that promotes motivation, development, and performance. A reliable and objective estimation of the flow is essential for moving away from the traditional self-reporting subjective questionnaires, and for developing closed-loop human-computer interfaces. In this study, we recorded EEG and pupil dilation in a cohort of participants solving arithmetic problems. In particular, the EEG activity was acquired with a prototype of a commercial headset from Logitech with nine dry electrodes incorporated in a pair of over-ear headphones. The difficulty of the tasks was adapted to induce mental *Boredom*, *Flow* and *Overload*, corresponding to too easy, optimal and too challenging tasks, respectively. Results indicated statistically significant differences between all pairs of conditions for the pupil dilation, as well as for the EEG activity for the electrodes in the ear-pads. Furthermore, we built a predictive model that estimated the mental state of the user from their EEG data with 65% accuracy.**

I. INTRODUCTION

In psychology, the flow state is considered to be the mental state in which a person is fully immersed and focused, feeling an intrinsic reward while engaging in a certain task [1]. Previous research showed that flow improves motivation [2], [3], development [4], [5] and performance [6], [7]. Understanding and estimating the flow experience with low-cost and non-invasive methods could facilitate the development of applications that can potentially evoke such mental state, thus optimizing the environment for users to learn and perform. For instance, such systems could prevent interruptions during work by ensuring that there are no distractions whenever the system detects the user is in the flow state [8].

Previous studies sought biomarkers of cognitive load and effort in pupil dilation [9], [10], [11], [12]. In an experiment in which participants were solving arithmetic tasks mentally, [13] showed that the pupil dilates gradually between the presentation of the problem and its solution, reaching its maximum before reporting the result and then returning to its original size. While mental effort and cognitive load might be related to mental flow, none of the above-mentioned pupillometry studies were designed to directly induce the flow state in participants.

Traditionally, the mental flow has been measured subjectively through questionnaires [14]. However, subjective self-reporting interrupts the activity and is thus not able to capture moment-to-moment variations. Previous research has attempted to measure the flow objectively, studying brain activity through Functional Magnetic Resonance Imaging (fMRI) [15] while completing mental arithmetic tasks of varying difficulty. However, the low temporal resolution of fMRI is problematic for rapidly detecting the flow state, and it is not feasible for everyday usage. To overcome these limitations, [16] replicated the experimental paradigm of mental arithmetic tasks from [15] but used research-grade electroencephalography (EEG) to seek neural correlates of mental flow.

EEG can be used to measure brain activity in cases where users might not be able to articulate their experience properly. Especially portable EEG setups do not interfere with a participant's behaviour [17], [18], and provide an objective source of information that does not rely on subjective reporting [19]. Analysing the EEG signal in different frequency bands (delta: 13 Hz, theta: 47 Hz, alpha: 813 Hz, beta: 1430 Hz) showed that the frontal theta-band power and alpha-band power in the frontal and the right-central areas increase as the participant enters the flow state [16]. The theta activity in the frontal area is related to cognitive load [20], [21] and concentration [22]. On the other hand, the increase in alpha activity is related to the working memory while performing the tasks [23], [21].

Typically, human neuroscience research is conducted in controlled laboratory environments. Thus, its contribution to understanding real-world scenarios is narrow, and the translation of findings outside the research lab is not trivial. In recent years, many non-invasive portable EEG systems have been developed [24]. Laboratory EEG headsets tend to have between 32 and 256 wet electrodes placed on all the areas of interest of the brain. However, commercial headsets usually have between 1 and 16 dry electrodes, which provide worse contact with the scalp and offer only sparse electrode montages, but at the same time improve the user experience [25]. Nevertheless, the portable EEG systems allow conducting "real-life" experiments in realistic environments (e.g., home or office) and continuously monitoring participants' brain activity over long periods of time.

Comparison of brain responses under different experimental conditions allows to investigate biomarkers of mental states. However, continuous estimation of mental states from EEG requires a dedicated decoding algorithm. Traditionally, EEG-based brain-computer interfaces (BCIs) consist of five steps [26]: EEG data acquisition; pre-processing and data cleaning; feature extraction; classification/estimation; and feedback to the user. Different paradigms within BCI require different signal processing [27], feature extraction [28] and

[1]Manuel Cherep and Pablo Mainar are with Logitech, Switzerland mcherep@logitech.com and pmainar@logitech.com
[2]Mikolaj Kegler is with Imperial College London, UK mikolaj.kegler16@imperial.ac.uk
[3]Jean-Philippe Thiran is a Full Professor at the École Polytechnique Fédérale de Lausanne, Switzerland jean-philippe.thiran@epfl.ch

classification methods [29]. Notably, the use of manually designed processing chain can be challenging and does not guarantee the optimal selection of features for a particular application.

The rise of deep learning has reduced the need for manual feature design while simultaneously achieving state-of-the-art performance [30]. In particular, Convolutional Neural Networks (CNNs) have been successful in many challenging problems, surpassing methods using domain knowledge for feature extraction [30], [31]. EEGNet [32] is a compact CNN that includes well-known EEG feature extraction concepts like optimal spatial filtering and filterbank construction, while at the same time reducing the size of the network compared to previous approaches [33] by two orders of magnitude. When evaluated on different EEG datasets from different BCI paradigms, EEGNet outperformed previous deep learning models in small datasets, as well as classic signal processing approaches. Table I shows the model architecture with all its parameters.

In this study, we sought to develop an algorithm for the mental flow estimation from wearable and portable EEG. To induce different mental states in the cohort of participants, we used the paradigm from [16], involving solving mental arithmetic tasks. During the task, we recorded participants' brain activity using a prototype of a commercial EEG headset with nine dry electrodes incorporated in a pair of over-ear headphones. In addition, we recorded their pupil dilation and asked them to complete subjective questionnaires without interrupting the activity, which served as a baseline for currently used methods for estimation of flow and cognitive load. We found significant differences between flow and all other mental states induced during the experiments in both EEG and pupil dilation. Our EEG biomarkers of mental flow, recorded using a wearable and portable system, were similar to [16], which used a research-grade EEG system. Finally, we designed a neural decoder for estimating participants' mental flow from EEG. Notably, using classic hand-crafted features yielded only chance level performance, while a deep learning-based model allowed to decode mental states with over 65% accuracy.

II. METHODOLOGY

A. Flow Experiments

The goal of these experiments was to collect data from participants performing tasks in different mental states, including flow, using the protocol of previous studies [15], [16]. The main difference between our experimental setup and that of previous research is that we used a prototype of a commercial Logitech EEG headset with nine dry electrodes and collected data using an eye tracker.

1) Participants: In this study, 11 Logitech employees (9 males, 2 females) participated in a total of 16 sessions where 5 participants completed the experiment twice on different days. We excluded three sessions from the dataset: one for covering the eye-tracker and two for poor contact of the EEG headset. The duration of the experiment was 1 hour, including the initial setup. All participants provided a written consent form following the declaration of Helsinki.

2) Experimental Design: The participants performed mental arithmetic tasks in different conditions corresponding to the level of difficulty. As they appeared on the screen, the tasks had to be solved mentally, and the results had to be submitted digit-by-digit with a mouse and a low-contrast on-screen keypad to limit their body movement. After submitting the result, there was no feedback regarding the correctness to avoid conditioning the participant.

The experiment had three conditions that correspond to different levels of task difficulty: Boredom (B), Flow (F) and Overload (O). The Boredom condition is characterized by tasks with a low level of difficulty, that provide no challenge to the participant. In the Flow condition, the difficulty of the tasks is adjusted dynamically to the participant's performance, providing a comfortable challenge that facilitates engagement. Finally, in the Overload condition, the difficulty is also dynamically adjusted to generate tasks that surpass their ability. Figure 1 summarizes visually the experimental design explained in detail below.

The Flow and Overload conditions only considered adding numbers of at most two digits, and used the same heuristic to increase or decrease the level of difficulty. To modify the difficulty, there are two scenarios: if the last summand of the last arithmetical expression had one digit, then the last summand in the next level changes to a two-digit number; and if it had two digits, then the next level adds a one-digit number at the end. Therefore, these changes occur alternatively. Decreasing the level follows the same logic in reverse order, going from two-digit to one-digit numbers or removing the last one-digit number. For example, level $1 \rightarrow$ "$12 + 76''$, level $2 \rightarrow$ "$23 + 86 + 2''$ and level 3 $\rightarrow$ "$64 + 28 + 55''$.

Given the above-mentioned heuristic, tasks in each condition were dynamically generated as follows:

- Boredom (B): the first summand is randomly generated from the range $[100, 109]$. The second summand is generated from the range $[1, 110 - \text{firstNumber}]$. This ensures that the total summation is always between 101 and 110.
- Flow (F): the adjustment of the difficulty is based on the results of the last two tasks. If the last two results are correct, then the level increases following the heuristic described above. On the other hand, if the last two results are incorrect, the level decreases following the same heuristic. Otherwise, the level remains the same. The level can decrease below the initial baseline, but it can never generate less than two summands of two-digit numbers.
- Overload (O): the level of difficulty increases when at least three out of the last five results are correct, and decreases if at least four out of the last five results are incorrect. Otherwise, the level remains the same. However, the level can never decrease below the initial baseline.

The conditions of Flow and Overload had an initial base-

TABLE I: EEGNet architecture and hyperparameters: C = number of EEG channels, T = input data frame size (samples), F_1 = number of temporal filters, D = depth multiplier (number of spatial filters), F_2 = number of pointwise filters, and N = number of classes. Table was reproduced based on [32].

Block	Layer	#filters	Size	#params	Output	Activation	Options
1	Input				(C, T)		
	Reshape				(1, C, T)		
	Conv2D	F_1	(1, 64)	$64 * F_1$	(F_1, C, T)	Linear	Mode = same
	BatchNorm			$2 * F_1$	(F_1, C, T)		
	DepthwiseConv2D	$D * F_1$	(C, 1)	$C * D * F_1$	($D * F_1$, 1, T)	Linear	Mode = valid, depth = D, max norm = 1
	BatchNorm			$2 * D * F_1$	($D * F_1$, 1, T)		
	Activation				($D * F_1$, 1, T)	ELU	
	AveragePool2D		(1, 4)		($D * F_1$, 1, T // 4)		
	Dropout*				($D * F_1$, 1, T // 4)		$p = 0.25$ or $p = 0.5$
2	SeparableConv2D	F_2	(1, 16)	$16 * D * F_1 + F_2 * (D * F_1)$	(F_2, 1, T // 4)	Linear	Mode = same
	BatchNorm			$2 * F_2$	(F_2, 1, T // 4)		
	Activation				(F_2, 1, T // 4)	ELU	
	AveragePool2D		(1, 8)		(F_2, 1, T // 32)		
	Dropout*				(F_2, 1, T // 32)		$p = 0.25$ or $p = 0.5$

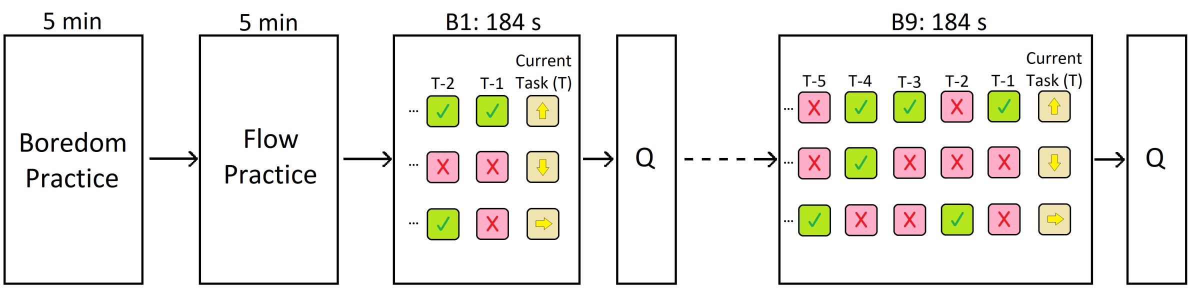


Fig. 1: Diagram of the experimental design: T stands for task (i.e. single arithmetic problem), B for block of tasks, and Q for questionnaire. Yellow arrows show how a task's level changes depending on previous results, where B1 would correspond to Flow and B9 to Overload.

line level that is set according to the skills of each participant. The Flow baseline was calculated based on a practice session before the experiment. The practice session starts with two summands of two-digit numbers and followed the same heuristics explained above to adjust the level for Flow. The baseline for the Flow condition was calculated as the average level of the last 25% of tasks in the practice session. However, to avoid a lower performance due to unfamiliarity with the interface, there was a first practice session with tasks corresponding to the Boredom condition, where participants could familiarize themselves with the experimental setup. Finally, the baseline for the Overload condition was set to three levels higher than the Flow baseline. These baselines reset for each block, independently of performance.

Each experiment session included three blocks of tasks for each condition, as well as three Rest (R) blocks. To eliminate the potentially confounding effect of block order, there were two block sequences: RBFOFROBOBFR or RBOFORFBFBOR, excluding the initial practice. One of which was selected randomly for each session. Each regular block lasted 184 seconds; both practice sessions lasted 5 minutes; individual tasks had a timeout of 18 seconds; there was a 4-second break between tasks, and rest blocks lasted 25 seconds. Participants were aware of the existence of a task timeout, but they never knew how much time was left. The total duration of the recording was approximately 45 minutes, excluding the initial setup.

3) Data Collection: We collected participants' EEG and pupil size during the experiments. Additionally, we also recorded information corresponding to the arithmetic tasks: initial timestamp, submission timestamp, the result of the task, answer of the participant, condition, and the number of digits in the expression. After each block of tasks, participants were asked to fill out the same electronic questionnaire used in [15] and [16]. We used the questionnaire only to assess participants' engagement, and we did not use that obtained data in our analyses.

The EEG data were recorded using a new commercial prototype device developed at Logitech. The device incorporated nine *Conscious Labs* dry electrodes [34] into a Logitech Pro X Wireless headset. There were 5 electrodes in the headband which approximately correspond to "C1", "C2", "C3", "C4", "Cz" in the International 10/20 EEG system. Two more recording electrodes were placed around the ear in each earpad of the headset. The reference and ground were located on the mastoids. The EEG data were recorded using the Cyton board from OpenBCI[1] and the signals were sampled at 125 Hz and streamed to the PC via low-latency Bluetooth. The eye-related data were recorded with a Tobii Pro Nano eye-tracker at 90 Hz in the same constant light conditions to ensure reliable measurements [35]. The captured data was synchronized via custom-written Python code using the Lab Streaming Layer (LSL) [36] protocol and LabRecorder [37].

4) Data Analysis: The performance in the experiment was measured as the percentage of correct answers and the average response time for each condition and participant. These results, as well as the ones from the biomarkers, were examined using the Games-Howell post-hoc test [38], a non-parametric approach designed to compare combinations of multiple conditions which does not assume homoscedasticity, normality, or equal sample sizes. Furthermore, Cohen's d [39] was chosen for calculating effect sizes. In both the pupil and EEG data, we decided to remove the first second of each task to account for the time that it took the participant to actually start solving the problem.

5) Pupil Dilation: The data collected from the eye-tracker includes the size of the pupil. However, we had to calculate the percentage of pupil dilation given the pupil size, to normalize the data across participants. The median pupil size, in the first 100 ms of each task, corresponded to the baseline used for calculating the evolution of the pupil dilation in time. The final pupil dilation examined in the statistical analysis was the average of both eyes.

The information from the eye-tracker contains blinks and other artifacts that had to be removed. To detect and remove these artifacts we defined our own method without parameters. We calculated the discrete derivative of the pupil size signal to see how much it varies from one point in time to the next one. Then, we applied a closing function from mathematical morphology, which yielded an envelope for the gradients, grouping together some of the outliers that couldn't be detected otherwise. Finally, we clipped the values below a threshold of 0.02, leaving only the artifacts to be removed.

The logic behind the threshold was that information-related dilations of the pupil rarely exceed .5 mm [40], [41]. The threshold in our method was constant and did not have to be fine-tuned for each individual, unlike most existing methods [42], [43]. After removing the artifacts, we smoothed the signal using a zero-phase low-pass first-order IIR filter with a 4 Hz cut-off frequency [44]. Finally, we excluded tasks with less than 40% of data remaining after the artifact removal.

6) EEG: For the pre-processing of the EEG data, we applied a notch filter at 25 and 50 Hz (i.e. power line noise and its subharmonic) as well as a band-pass IIR first-order filter between 1 and 50 Hz.

Afterwards, we calculated the power spectral density (PSD) for each of the four bands, in each condition, with epochs of 1 second. During epoching, we rejected epochs exceeding 2 standard deviations of the peak-to-peak signal amplitude (PTP) for each channel. The PSD in different conditions and from different sensors, were compared in the statistical analysis.

[1]https://openbci.com/

B. EEGNet

EEGNet [32] is the model that we used to classify raw EEG into our three classes: Boredom, Flow and Overload; in order to predict the mental state from single-trial (i.e. one arithmetic task) EEG. We chose $F_1 = 8, D = 2, F_2 = 16, C = 9, T = 126, N = 3, p = 0.5$ as well as kernel length $= 64$ and batch $= 64$ as training parameters (see Table I), following the same pre-processing presented in Section II-A.6 as well as standardization by channel for each individual. The channel-wise mean and standard deviation were computed from the training data only.

The original dataset was divided into training (80%) and validation (20%) sets, where each participant contributed those percentages of data to the collective final dataset. Given the characteristics of the experiment, different levels of difficulty led to different response times, which caused data imbalance across conditions. In the Boredom condition, tasks were easy and took considerably less time than those in the Flow and Overload conditions. Therefore, in terms of seconds of data for EEG, Boredom was under-represented, as compared to other conditions. Considering this data imbalance, we trained with weights according to each class. The model was trained for 5000 epochs using Adam [45] and the categorical cross-entropy loss. For comparison, we used a Random Forest [46] classifier (max. tree depth: 5, number of trees: 1000, min. samples leaf: 1) as a baseline with the same pre-processing, taking the power bands from the electrodes in the ear-pads instead of the raw data.

C. Implementation Details

All of the above methods were implemented using custom-written Python code using the following open-source packages: MNE [47], Scipy [48], pandas [49], NumPy [50], Pingouin [51], pyXDF [52], Keras [53], Scikit-Learn [54], Scikit-Image [55] and imbalanced-learn [56].

III. RESULTS

A. Behavioral Data

The performance of the mental arithmetic tasks was measured in terms of the percentage of correct answers. The average performance $\pm$ 1 standard deviation in the Boredom condition was $99.04 \pm 2.54\%$, in the Flow condition $46.04 \pm 17.34\%$, and in the Overload condition $15.32 \pm 16.15\%$. The average time response was 2.70 ± 0.37 seconds in Boredom, 14.37 ± 2.31 seconds in Flow and 17.54 ± 0.9 seconds in Overload. These results are in agreement with previous studies [16], [15], and indicate a successful induction of mental states in the participants.

B. Physiological Data

Figure 2a presents the distribution of pupil dilation per condition considering all participants. As we can observe, the pupil dilated the least in Boredom, followed by Overload and yielded the highest average dilation during Flow. There were statistically significant ($p < 0.05$) differences in pupil dilation between all pairs of conditions, considering the data for all participants (see Table II), but the associated effect sizes are small (i.e. below 0.3 Cohen's d).

Figure 2b presents the distribution of EEG power band per condition, considering all participants. The Flow condition yielded the highest power and could be distinguished from other states the easiest. In particular, we found statistically significant differences ($p < 0.05$) in EEG band power between Boredom and Flow as well as between Flow and Overload in theta, alpha and beta for the electrodes in the ear-pads (see Table II). The effect sizes were comparable to those obtained for pupil dilation.

TABLE II: Global effect size (Cohen's d) for the multiple comparisons in pupil and EEG, including different frequency bands and channel subsets. Bold values correspond to a p-value < 0.05.

Feature	Electrodes	Boredom/Flow	Boredom/Overload	Flow/Overload
Pupillometry	-	**-0.286**	**-0.121**	**0.140**
EEG-Delta	All	0.004	-0.032	-0.037
EEG-Theta	All	-0.022	-0.030	-0.010
EEG-Alpha	All	**-0.077**	0.043	**0.126**
EEG-Beta	All	-0.062	**0.114**	**0.194**
EEG-Delta	Headband	-0.011	-0.071	**-0.062**
EEG-Theta	Headband	-0.028	-0.076	-0.050
EEG-Alpha	Headband	-0.071	0.006	**0.077**
EEG-Beta	Headband	-0.040	**0.098**	**0.141**
EEG-Delta	Ear-pads	-0.071	-0.031	0.042
EEG-Theta	Ear-pads	**-0.108**	-0.014	**0.100**
EEG-Alpha	Ear-pads	**-0.161**	0.066	**0.257**
EEG-Beta	Ear-pads	**-0.161**	**0.092**	**0.319**

Figure 3 presents the validation accuracies of models for predicting the mental state from single-trial EEG data. The average validation accuracy for EEGNet was 65%, and 43% for the Random Forest. Moreover, all participants except one achieved higher accuracies for EEGNet. In particular, Table III shows the precision, recall and F1 score for each class of EEGNet. The obtained scores were considerably lower in the Boredom condition, as compared to Flow and Overload, which were similar. This implies that Boredom was misclassified more often than other conditions.

TABLE III: Precision, recall and F1 for the EEGNet model.

Mental state	Precision	Recall	F1-Score
Boredom	0.41	0.43	0.42
Flow	0.66	0.70	0.68
Overload	0.71	0.66	0.68

IV. DISCUSSION

The goal of this study was to identify biomarkers corresponding to the flow state using mental arithmetic tasks designed to induce the mental flow state, as well as to build a model that could continuously estimate the user's mental state. In particular, all the experiments were conducted with a prototype of a commercial EEG headset. Therefore our results contribute to understanding whether biomarkers of flow could be estimated outside research laboratories without access to research-grade equipment.

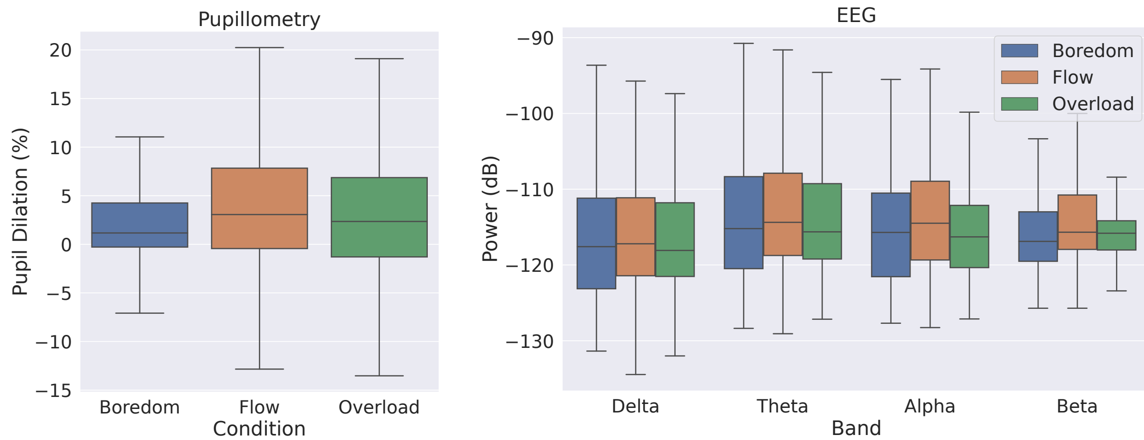
In conclusion, high theta activity, related to high cognitive load, and moderate alpha activity, related to low working memory, were detected in the electrodes in the ear-pads, suggesting that a combination of both might represent the state of flow. Overall, these results agree with previous findings from [16], which used research-grade EEG. The largest significant effects were obtained for the electrodes in the ear-pads, which might be explained by the better contact of these electrodes with the scalp. In addition, we showed that the EEG power yields similar trends and effects to pupil dilation measured using a commercial-grade eye tracker.

Furthermore, we built a model for decoding mental states from single-trial EEG data. We compared a Random Forest classifier using EEG band power features commonly used in BCI, with the light-weight EEGNet. The Random Forest classifier using hand-crafted features failed to predict the mental states and achieved only 43%, only slightly exceeding the chance level (33%). EEGNet, on the other hand, achieved a significantly higher accuracy of 65%. While the EEG band power features were significantly different for all conditions, the effect sizes were considerably low. This implied rather small trial-to-trial differences, which were not consistent enough to yield good single-trial classification using simple EEG band power features. The EEGNet model was able to learn to extract optimal features from EEG, which led to the improved single-trial mental state classification accuracy.

Although our results indicate that the wearable EEG system embedded in the conventional headset can detect biomarkers of mental flow, the study had several limitations. Firstly, the experimental paradigm proposed by [16] is still distant from a real-life scenario. To validate the proposed approach, it should be applied to the long-term continuous recording of brain activity during realistic tasks, such as office work, coding, etc. Unlike research-grade EEG, which does not offer much mobility, our headset prototype allows the user to move freely, which makes such a follow-up study possible.

Secondly, we were able to recruit only several participants to take part in the experiment twice. While the mental state decoding was generally stable between the two sessions (Figure 3, subjects with _2 suffixes), we haven't extensively studied the longitudinal stability of the decoder. The use of dry electrodes and misalignment of the headset might increase between-session variability, which should be investigated. Future work should systematically study subject-independent performance of the proposed approach.

Thirdly, the experimental design introduced in [16] could be improved. The biggest problem we encountered was the data imbalance, especially in the Boredom condition associated with easy tasks. In this condition, participants



(a) Percentage of pupil dilation.

(b) EEG band power for electrodes in the ear-pads.

Fig. 2: Distribution of data per condition.

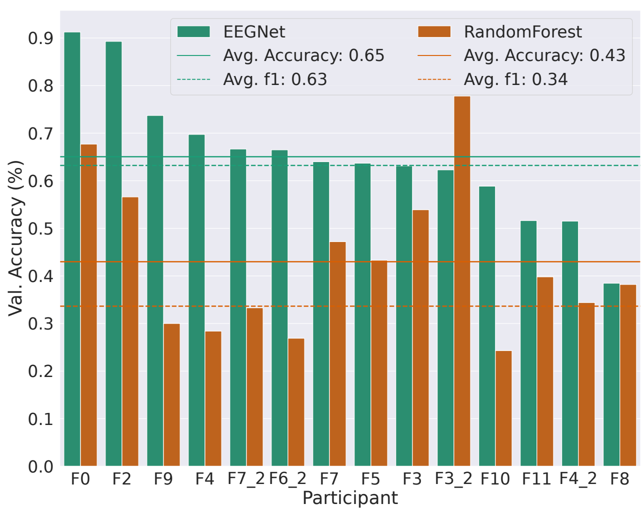


Fig. 3: Single-trial mental state decoding validation accuracy. IDs of participants who attended the second recording session are suffixed with _2.

solved the problems quickly, which resulted in less task-related EEG data. This, in turn, impacted the ability of our models to reliably detect this mental state. The experimental setup could be improved by re-designing the task to sustain the induced mental state for longer. This would also solve the fact that most of the time spent in the Boredom condition was mouse movement and not mental calculation. Moreover, most participants reported that some Overload tasks seemed too difficult and therefore gave up.

Finally, while the presented results are promising for using wearable EEG for estimating mental states, the decoding accuracy could be improved. In particular, the continuously predicted mental states could benefit from an additional probabilistic model applied to the output of the EEGNet, which would provide long-term context by integrating consecutive predictions. Furthermore, the model could benefit from including other biosignals, such as heart or respiratory rate, which were previously used to predict mental flow [57].

V. Conclusions

We used wearable EEG embedded in a pair of over-ear headphones to estimate the user's mental flow during mental arithmetic tasks. We found significant differences in EEG band power during the flow states and other experimental conditions inducing boredom or mental overload, and the effect size was comparable to that obtained from pupillometry, often used for estimating cognitive load. We used the deep learning model EEGNet to estimate mental states from short periods of EEG data corresponding to single trials of solving arithmetic problems. The proposed model achieved 65% accuracy in the 3-way classification problem, while the conventional classifier based on the EEG band power features only slightly exceeded the chance level. Our results illustrate the potential of using wearable EEG for continuous estimation of the mental flow state in real-life scenarios outside the controlled research environment.

Acknowledgments

We thank Aindrias Lefévère-Laoide for help with the data collection.

References

[1] Csikszentmihalyi, M. Beyond boredom and anxiety. (Jossey-Bass, 2000)

[2] Csikszentmihalyi, M., Rathunde, K. & Whalen, S. Talented teenagers: The roots of success and failure. (Cambridge University Press, 1997)

[3] Jackson, S., Ford, S., Kimiecik, J. & Marsh, H. Psychological correlates of flow in sport. *Journal Of Sport And Exercise Psychology*. **20**, 358-378 (1998)

[4] Carli, M., Fave, A. & Massimini, F. The quality of experience in the flow channels: Comparison of Italian and US students. (Cambridge University Press, 1988)

[5] Nakamura, J. Optimal experience and the uses of talent. (Cambridge University Press, 1988)

[6] Jackson, S., Thomas, P., Marsh, H. & Smethurst, C. Relationships between flow, self-concept, psychological skills, and performance. *Journal Of Applied Sport Psychology*. **13**, 129-153 (2001)

[7] MacDonald, R., Byrne, C. & Carlton, L. Creativity and flow in musical composition: An empirical investigation. *Psychology Of Music*. **34**, 292-306 (2006)

[8] Rissler, R., Nadj, M., Li, M., Knierim, M. & Maedche, A. Got flow? Using machine learning on physiological data to classify flow. *Extended Abstracts Of The 2018 CHI Conference On Human Factors In Computing Systems*. pp. 1-6 (2018)

[9] Bradley, M., Miccoli, L., Escrig, M. & Lang, P. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*. **45**, 602-607 (2008)

[10] Chapman, C., Oka, S., Bradshaw, D., Jacobson, R. & Donaldson, G. Phasic pupil dilation response to noxious stimulation in normal volunteers: relationship to brain evoked potentials and pain report. *Psychophysiology*. **36**, 44-52 (1999)

[11] Kahneman, D. & Beatty, J. Pupil diameter and load on memory. *Science*. **154**, 1583-1585 (1966)

[12] Hoeks, B. & Levelt, W. Pupillary dilation as a measure of attention: A quantitative system analysis. *Behavior Research Methods, Instruments, & Computers*. **25**, 16-26 (1993)

[13] Hess, E. & Polt, J. Pupil size in relation to mental activity during simple problem-solving. *Science*. **143**, 1190-1192 (1964)

[14] Moneta, G. On the measurement and conceptualization of flow. *Advances In Flow Research*. pp. 23-50 (2012)

[15] Ulrich, M., Keller, J., Hoenig, K., Waller, C. & Grön, G. Neural correlates of experimentally induced flow experiences. *Neuroimage*. **86** pp. 194-202 (2014)

[16] Katahira, K., Yamazaki, Y., Yamaoka, C., Ozaki, H., Nakagawa, S. & Nagata, N. EEG correlates of the flow state: A combination of increased frontal theta and moderate frontocentral alpha rhythm in the mental arithmetic task. *Frontiers In Psychology*. **9** pp. 300 (2018)

[17] Mayer, R. How can brain research inform academic learning and instruction?. *Educational Psychology Review*. **29**, 835-846 (2017)

[18] Hölle, D., Meekes, J. & Bleichner, M. Mobile ear-EEG to study auditory attention in everyday life. *Behavior Research Methods*. pp. 1-12 (2021)

[19] Dahlstrom-Hakki, I., Asbell-Clarke, J. & Rowe, E. Showing is knowing: The potential and challenges of using neurocognitive measures of implicit learning in the classroom. *Mind, Brain, And Education*. **13**, 30-40 (2019)

[20] Cavanagh, J. & Frank, M. Frontal theta as a mechanism for cognitive control. *Trends In Cognitive Sciences*. **18**, 414-421 (2014)

[21] Klimesch, W. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews*. **29**, 169-195 (1999)

[22] Lagopoulos, J., Xu, J., Rasmussen, I., Vik, A., Malhi, G., Eliassen, C., Arntsen, I., Sæther, J., Hollup, S., Holen, A. & Others Increased theta and alpha EEG activity during nondirective meditation. *The Journal Of Alternative And Complementary Medicine*. **15**, 1187-1192 (2009)

[23] Tuladhar, A., Huurne, N., Schoffelen, J., Maris, E., Oostenveld, R. & Jensen, O. Parieto-occipital sources account for the increase in alpha activity with working memory load. *Human Brain Mapping*. **28**, 785-792 (2007)

[24] LaRocco, J., Le, M. & Paeng, D. A systemic review of available low-cost EEG headsets used for drowsiness detection. *Frontiers In Neuroinformatics*. **14** (2020)

[25] Hinrichs, H., Scholz, M., Baum, A., Kam, J., Knight, R. & Heinze, H. Comparison between a wireless dry electrode EEG system with a conventional wired wet electrode EEG system for clinical applications. *Scientific Reports*. **10**, 1-14 (2020)

[26] Nicolas-Alonso, L. & Gomez-Gil, J. Brain computer interfaces, a review. *Sensors*. **12**, 1211-1279 (2012)

[27] Bashashati, A., Fatourechi, M., Ward, R. & Birch, G. A survey of signal processing algorithms in braincomputer interfaces based on electrical brain signals. *Journal Of Neural Engineering*. **4**, R32 (2007)

[28] McFarland, D., Anderson, C., Muller, K., Schlogl, A. & Krusienski, D. BCI meeting 2005-workshop on BCI signal processing: feature extraction and translation. *IEEE Transactions On Neural Systems And Rehabilitation Engineering*. **14**, 135-138 (2006)

[29] Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F. & Arnaldi, B. A review of classification algorithms for EEG-based braincomputer interfaces. *Journal Of Neural Engineering*. **4**, R1 (2007)

[30] LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature*. **521**, 436-444 (2015)

[31] Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Networks*. **61** pp. 85-117 (2015)

[32] Lawhern, V., Solon, A., Waytowich, N., Gordon, S., Hung, C. & Lance, B. EEGNet: a compact convolutional neural network for EEG-based braincomputer interfaces. *Journal Of Neural Engineering*. **15**, 056013 (2018)

[33] Schirrmeister, R., Springenberg, J., Fiederer, L., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W. & Ball, T. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping*. **38**, 5391-5420 (2017)

[34] Armand, M., Zou, S. & Dauguet, J. Patent FR3089991: Composition Polymérique Conductrice. (2020)

[35] Kang, O., Huffer, K. & Wheatley, T. Pupil dilation dynamics track attention to high-level information. *PloS One*. **9**, e102463 (2014)

[36] Kothe, C., Medine, D., Boulay, C., Grivich, M. & Stenner, T. LSL: a system for the unified collection of measurement time series in research experiments.

[37] Stenner, T., Boulay, C. & Medine, D. LabRecorder: An application for recording LSL streams in XDF file format.. (https://github.com/labstreaminglayer/App-LabRecorder,2021)

[38] Games, P. & Howell, J. Pairwise multiple comparison procedures with unequal ns and/or variances: a Monte Carlo study. *Journal Of Educational Statistics*. **1**, 113-125 (1976)

[39] Cohen, J. Statistical power analysis for the behavioral sciences. (Academic press, 1969)

[40] Beatty, J., Lucero-Wagoner, B. & Others The pupillary system. *Handbook Of Psychophysiology*. **2** (2000)

[41] Goldwater, B. Psychological significance of pupillary movements.. *Psychological Bulletin*. **77**, 340 (1972)

[42] Mathôt, S., Fabius, J., Van Heusden, E. & Stigchel, S. Safe and sensible preprocessing and baseline correction of pupil-size data. *Behavior Research Methods*. **50**, 94-106 (2018)

[43] Kret, M. & Sjak-Shie, E. Preprocessing pupil size data: Guidelines and code. *Behavior Research Methods*. **51**, 1336-1342 (2019)

[44] Jackson, I. & Sirois, S. Infant cognition: going full factorial with pupil dilation. *Developmental Science*. **12**, 670-679 (2009)

[45] Kingma, D. & Ba, J. ADAM: A Method for Stochastic Optimization. (2014)

[46] Breiman, L. Random forests. *Machine Learning*. **45**, 5-32 (2001)

[47] Gramfort, A., Luessi, M., Larson, E., Engemann, D., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L. & Hämäläinen, M. MEG and EEG Data Analysis with MNE-Python. *Frontiers In Neuroscience*. **7**, 1-13 (2013)

[48] Virtanen, P., Gommers, R., Oliphant, T., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Van der Walt, S., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A., Jones, E., Kern, R., Larson, E., Carey, C., Polat, ., Feng, Y., Moore, E., Vand erPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E., Harris, C., Archibald, A., Ribeiro, A., Pedregosa, F., Van Mulbregt, P. & Contributors, S. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*. **17** pp. 261-272 (2020)

[49] McKinney Data Structures for Statistical Computing in Python. *Proceedings Of The 9th Python In Science Conference*. pp. 56 - 61 (2010)

[50] Oliphant, T. A guide to NumPy. (Trelgol Publishing USA,2006)

[51] Vallat, R. Pingouin: statistics in Python. *Journal Of Open Source Software*. **3**, 1026 (2018)

[52] Boulay, C., Brunner, C. & Stenner, T. pyXDF: a Python importer for XDF files.. (https://github.com/xdf-modules/pyxdf, 2020)

[53] Chollet, F. & Others Keras. (GitHub, 2015), https://github.com/fchollet/keras

[54] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, E. Scikit-learn: Machine Learning in Python. *Journal Of Machine Learning Research*. **12** pp. 2825-2830 (2011)

[55] Walt, S., Schönberger, J., Nunez-Iglesias, J., Boulogne, F., Warner, J., Yager, N., Gouillart, E., Yu, T. & Contributors scikit-image: image processing in Python. *PeerJ*. **2** pp. e453 (2014,6)

[56] Lemaître, G., Nogueira, F. & Aridas, C. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *Journal Of Machine Learning Research*. **18**, 1-5 (2017), http://jmlr.org/papers/v18/16-365.html

[57] Knierim, M., Rissler, R., Dorner, V., Maedche, A. & Weinhardt, C. The psychophysiology of flow: a systematic review of peripheral nervous system features. *Information Systems And Neuroscience*. pp. 109-120 (2018)