DATASETS FOR BRAIN-COMPUTER INTERFACE APPLICATIONS

EDITED BY: Ian Daly, Ana Matran-Fernandez, Davide Valeriani, Mikhail Lebedev and Andrea Kübler

PUBLISHED IN: Frontiers in Neuroscience

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88971-694-4
DOI 10.3389/978-2-88971-694-4

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

DATASETS FOR BRAIN-COMPUTER INTERFACE APPLICATIONS

Topic Editors:
Ian Daly, University of Essex, United Kingdom
Ana Matran-Fernandez, University of Essex, United Kingdom
Davide Valeriani, Neurable Inc., United States
Mikhail Lebedev, Skolkovo Institute of Science and Technology, Russia
Andrea Kübler, Julius Maximilian University of Würzburg, Germany

Citation: Daly, I., Matran-Fernandez, A., Valeriani, D., Lebedev, M., Kübler, A., eds. (2021). Datasets for Brain-Computer Interface Applications. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-694-4

Table of Contents

frontiers in Neuroscience

EDITORIAL
published: 29 September 2021
doi: 10.3389/fnins.2021.732165

Editorial: Datasets for Brain-Computer Interface Applications

Ian Daly[1]*, *Ana Matran-Fernandez*[1], *Davide Valeriani*[2], *Mikhail Lebedev*[3,4] and *Andrea Kübler*[5]

[1] Brain-Computer Interfacing and Neural Engineering Laboratory, School of Computer Science and Electronic Engineering, University of Essex, Colchester, United Kingdom, [2] Neurable Inc., Boston, MA, United States, [3] V. Zelman Center for Neurobiology and Brain Restoration, Skolkovo Institute of Science and Technology, Moscow, Russia, [4] Department of Information and Internet Technologies of Digital Health Institute, I.M. Sechenov First Moscow State Medical University, Moscow, Russia, [5] Institute of Psychology, Julius Maximilian University of Würzburg, Würzburg, Germany

Keywords: brain computer interface, datasets, EEG, functional near infrared spectroscopy, electromyogram, ERP, steady-state visual evoked potential, affect (emotion)

Editorial on the Research Topic

Datasets for Brain-Computer Interface Applications

Non-invasive Brain-computer interfaces are an exciting new technology that provide a channel for communication between the brain and a computer system. They can be used as communication devices (Chaudhary et al., 2016; Brumberg et al., 2018), rehabilitation systems (Cervera et al., 2018), entertainment devices (Gürkök et al., 2017), and for a wide range of other applications (Finke et al., 2009; Makeig et al., 2011).

Research in non-invasive BCIs is developing rapidly and is a highly multidisciplinary field, involving, among others, neuroscientists, engineers, psychologists, computer scientists, and clinicians. Continuing development of BCI technology relies on advances made in each of these fields, which individually and collectively can contribute to improving all aspects of BCI systems including signal acquisition, processing, classification, and user interface design.

Many individual parts of a BCI system are typically first developed and evaluated on pre-existing datasets. However, there are only a few high quality publicly available datasets on which new systems, tools, and technologies can be evaluated and compared. For example, the publicly available BCI competition datasets (Sajda et al., 2003; Blankertz et al., 2004, 2006) provide an excellent set of resources for BCI researchers and have been widely used by numerous researchers to develop and evaluate new signal processing and classification methods (Arvaneh et al., 2013; Ghaemi et al., 2017; Lotte et al., 2018; Sakhavi et al., 2018; Zanini et al., 2018; Zhang et al., 2018). Yet, the relatively small size and number of such datasets introduce the risk of overfitting to methods developed and evaluated with these datasets. In other words, the reliability and reproducibility of BCI research is held back by a lack and sparsity of publicly available datasets.

This special issue provides a collection of descriptions of publicly available physiological datasets recorded during development, training, and evaluation of non-invasive BCI systems from BCI research labs around the world.

The collected datasets consist of signals recorded via a wide variety of modalities, including, but not limited to, electroencephalography (EEG), functional near infrared spectroscopy (fNIRS), electromyography (EMG), electrocardiography (ECG), galvanic skin response (GSR), skin temperature measures, respiration rates, and body movement data. Many datasets include multi-modal recordings with combinations of two or more of these signal modalities.

Data from a wide variety of different BCI paradigms are described. These include development of novel event-related potential (ERP) and steady state visual evoked potential (SSVEP) based BCIs

OPEN ACCESS

Edited and reviewed by:
Michele Giugliano,
International School for Advanced Studies (SISSA), Italy

***Correspondence:**
Ian Daly
i.daly@essex.ac.uk

Specialty section:
This article was submitted to Neural Technology, a section of the journal Frontiers in Neuroscience

Received: *28 June 2021*
Accepted: *12 August 2021*
Published: *29 September 2021*

Citation:
Daly I, Matran-Fernandez A, Valeriani D, Lebedev M and Kübler A (2021) Editorial: Datasets for Brain-Computer Interface Applications. Front. Neurosci. 15:732165. doi: 10.3389/fnins.2021.732165

for communication, motor imagery BCIs, affective BCIs, collaborative BCIs, and neurofeedback-based BCIs for nicotine addiction, as well as resting-state data.

Data on ERP-based BCIs are provided by several authors. For example, Delijorge et al. describe an EEG-based P300-based robotic hand control BCI; Simões et al. provide a large EEG-based P300-based BCI dataset; Li et al. implemented an ERP-based BCI for communication.

Motor control-based BCIs and associated datasets are also included in this collection. For example, Brandl and Blankertz provide an EEG dataset recorded during motor imagery while distractions were presented to simulate day-to-day events experienced outside the lab. Schwarz et al. made an attempt to decode reach and grasp actions from the EEG. Ortega et al. collected a multimodal dataset comprising EEG, fNIRS, EMG, and movement data recorded during a force grip task.

A wide range of other types of EEG-based BCIs are also presented. These include a dataset for a BCI based on covert attention shifts (Reichert et al.) and an affective BCI based on neurofeedback (Charles et al.), as well as two BCIs based on the rapid serial visual presentation paradigm (Zhang et al.; Zheng et al.). The collection also includes a BCI for treating nicotine addiction via neurofeedback (Bu et al.) and a dataset of SSVEP signals (Liu et al.).

A diverse range of paradigms were used in this collection of studies. For example, von Lühmann et al. present a resting state fNIRS dataset, while Parent et al. provide a multimodal dataset, comprising EEG, ECG, and respiration activity, recorded during a range of physical activities and induced stress. Finally, Albuquerque et al. offer a multimodal dataset, comprising EEG, ECG, and GSR, recorded during a mental workload paradigm.

We expect that the collected datasets will enable novel developments and applications of BCI technology, as well as extensive validation studies of current and future BCIs.

AUTHOR CONTRIBUTIONS

All authors co-wrote the editorial and edited the Research Topic.

FUNDING

ML was supported by the Russian Science Foundation grant 21-75-30024.

REFERENCES

Arvaneh, M., Guan, C., Ang, K. K., and Quek, C. (2013). Optimizing spatial filters by minimizing within-class dissimilarities in electroencephalogram-based brain-computer interface. *IEEE Transac. Neural Netw. Learn. Syst.* 24, 610–619. doi: 10.1109/TNNLS.2013.2239310

Blankertz, B., Müller, K.-R., Curio, G., Vaughan, T. M., Schalk, G., Wolpaw, J. R., et al. (2004). The BCI Competition 2003: progress and perspectives in detection and discrimination of EEG single trials. *IEEE Trans. Biomed. Eng.* 51, 1044–1051. doi: 10.1109/TBME.2004.826692

Blankertz, B., Müller, K.-R., Krusienski, D. J., Schalk, G., Wolpaw, J. R., Schlögl, A., et al. (2006). The BCI competition. III: validating alternative approaches to actual BCI problems. *IEEE Transac. Neural Syst. Rehabil. Eng.* 14, 153–159. doi: 10.1109/TNSRE.2006.875642

Brumberg, J. S., Pitt, K. M., Mantie-Kozlowski, A., and Burnison, J. D. (2018). Brain–computer interfaces for augmentative and alternative communication: a tutorial. *Am. J. Speech Lang. Pathol.* 27, 1–12. doi: 10.1044/2017_AJSLP-16-0244

Cervera, M. A., Soekadar, S. R., Ushiba, J., Millán, J. D. R., Liu, M., Birbaumer, N., et al. (2018). Brain-computer interfaces for post-stroke motor rehabilitation: a meta-analysis. *Ann. Clin. Transl. Neurol.* 5, 651–663. doi: 10.1002/acn3.544

Chaudhary, U., Birbaumer, N., and Ramos-Murguialday, A. (2016). Brain-computer interfaces for communication and rehabilitation. *Nat. Rev. Neurol.* 12, 513–525. doi: 10.1038/nrneurol.2016.113

Finke, A., Lenhardt, A., and Ritter, H. (2009). The MindGame: a P300-based brain-computer interface game. *Neural Netw.* 22, 1329–1333. doi: 10.1016/j.neunet.2009.07.003

Ghaemi, A., Rashedi, E., Pourrahimi, A. M., Kamandar, M., and Rahdari, F. (2017). Automatic channel selection in EEG signals for classification of left or right hand movement in Brain Computer Interfaces using improved binary gravitation search algorithm. *Biomed. Signal Process. Control.* 33, 109–118. doi: 10.1016/j.bspc.2016.11.018

Gürkök, H., Hakvoort, G., Poel, M., and Nijholt, A. (2017). Meeting the expectations from brain-computer interfaces. *Comput. Entertain.* 15:5. doi: 10.1145/2633431

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., et al. (2018). A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update. *J. Neural Eng.* 15:031005. doi: 10.1088/1741-2552/aab2f2

Makeig, S., Leslie, G., Mullen, T., Sarma, D., Bigdely-Shamlo, N., and Kothe, C. (2011). "First demonstration of a musical emotion BCI," in *International Conference on Affective Computing and Intelligent Interaction* (Berlin; Heidelberg: Springer) 487–496.

Sajda, P., Gerson, A., Müller, K.-R. K.-R., Blankertz, B., and Parra, L. (2003). A data analysis competition to evaluate machine learning algorithms for use in brain-computer interfaces. *IEEE Transac. Neural Syst. Rehabil. Eng.* 11, 184–185. doi: 10.1109/TNSRE.2003.814453

Sakhavi, S., Guan, C., and Yan, S. (2018). Learning Temporal Information for Brain-Computer Interface Using Convolutional Neural Networks. *IEEE Transac. Neural Netw. Learn. Syst.* 29, 5619–5629. doi: 10.1109/TNNLS.2018.2789927

Zanini, P., Congedo, M., Jutten, C., Said, S., and Berthoumieu, Y. (2018). Transfer learning: a riemannian geometry framework with applications to brain-computer interfaces. *IEEE Transac. Biomed. Eng.* 65, 1107–1116. doi: 10.1109/TBME.2017.2742541

Zhang, Y., Wang, Y., Zhou, G., Jin, J., Wang, B., Wang, X., et al. (2018). Multi-kernel extreme learning machine for EEG classification in brain-computer interfaces. *Expert Syst. Appl.* 96, 302–310. doi: 10.1016/j.eswa.2017.12.015

Conflict of Interest: DV is employed by Neurable Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 31 January 2020
doi: 10.3389/fnins.2020.00054

Comparison of the ERP-Based BCI Performance Among Chromatic (RGB) Semitransparent Face Patterns

Shurui Li[1], Jing Jin[1], Ian Daly[2], Cili Zuo[1], Xingyu Wang[1] and Andrzej Cichocki[3,4,5]*

[1] Key Laboratory of Advanced Control and Optimization for Chemical Processes, Ministry of Education, East China University of Science and Technology, Shanghai, China, [2] Brain-Computer Interfacing and Neural Engineering Laboratory, School of Computer Science and Electronic Engineering, University of Essex, Colchester, United Kingdom, [3] Skolkowo Institute of Science and Technology, Moscow, Russia, [4] Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland, [5] Department of Informatics, Nicolaus Copernicus University, Toruń, Poland

Objective: Previous studies have shown that combing with color properties may be used as part of the display presented to BCI users in order to improve performance. Build on this, we explored the effects of combinations of face stimuli with three primary colors (RGB) on BCI performance which is assessed by classification accuracy and information transfer rate (ITR). Furthermore, we analyzed the waveforms of three patterns.

Methods: We compared three patterns in which semitransparent face is overlaid three primary colors as stimuli: red semitransparent face (RSF), green semitransparent face (GSF), and blue semitransparent face (BSF). Bayesian linear discriminant analysis (BLDA) was used to construct the individual classifier model. In addition, a Repeated-measures ANOVA (RM-ANOVA) and Bonferroni correction were chosen for statistical analysis.

Results: The results indicated that the RSF pattern achieved the highest online averaged accuracy with 93.89%, followed by the GSF pattern with 87.78%, while the lowest performance was caused by the BSF pattern with an accuracy of 81.39%. Furthermore, significant differences in classification accuracy and ITR were found between RSF and GSF ($p < 0.05$) and between RSF and BSF patterns ($p < 0.05$).

Conclusion: The semitransparent faces colored red (RSF) pattern yielded the best performance of the three patterns. The proposed patterns based on ERP-BCI system have a clinically significant impact by increasing communication speed and accuracy of the P300-speller for patients with severe motor impairment.

Keywords: brain-computer interface, ERP, chromatic stimuli, semitransparent face, visual stimuli

OPEN ACCESS

Edited by:
Hans-Eckhardt Schaefer, University of Stuttgart, Germany

Reviewed by:
Zhen Liang, Kyoto University, Japan
Minpeng Xu, Tianjin University, China

****Correspondence:***
Jing Jin
jinjingat@gmail.com

Specialty section:
This article was submitted to Neural Technology, a section of the journal Frontiers in Neuroscience

Received: *25 September 2019*
Accepted: *14 January 2020*
Published: *31 January 2020*

Citation:
Li S, Jin J, Daly I, Zuo C, Wang X and Cichocki A (2020) Comparison of the ERP-Based BCI Performance Among Chromatic (RGB) Semitransparent Face Patterns. Front. Neurosci. 14:54. doi: 10.3389/fnins.2020.00054

INTRODUCTION

Brain-computer interface (BCI) systems enable their users to achieve direct communication with others or the outside environment by brain activity alone, independent of muscle control. There are many potential user groups for BCI systems, including, but not limited to, individuals living with amyotrophic lateral sclerosis (ALS) who are in the locked-in state (LIS).

The brain activity used to control a BCI can be measured using different signal acquisition approaches such as electroencephalogram (EEG), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), electrocorticogram (ECoG), or near infrared spectroscopy (NIRS) (Vidal, 1973, 1977; Wolpaw et al., 2002). Since EEG signals are recorded via non-invasive electrodes placed on the surface of the scalp, EEG-based BCI systems are very commonly used. Three key signal components of the EEG are frequently used for BCI control: event-related potentials (ERPs), steady-state visual evoked potentials (SSVEP), and motor imagery (MI) (Sutton et al., 1965; Coles and Rugg, 1995). The focus of the present study is the ERP-based BCI.

The P300 speller, is a visual ERP-based BCI system, that can elicit a P300 ERP component using an Oddball paradigm. The P300 potential is the largest positive deflection with a latency around 300 ms after the oddball stimulus onset, and is associated with various cognitive processes such as attention, working memory, and executive function (Van Dinteren et al., 2014). In addition, P300-based BCI systems can evoke P100, N200, and N400 components. The P300 speller was originally described by Farwell and Donchin (1988). In this study, participants were requested to watch a screen displaying a 6 × 6 matrix containing 26 letters and 10 digits. They were asked to focus on the rare target stimuli and ignore the common non-target stimuli. Stimuli were flashed (highlighted) in a row-column pattern (RCP). However, the RCP results in the adjacency-distraction and double-flash problems, which can cause false positive P300 ERPs during flashes of non-target stimuli that are adjacent to the target. Thus, some researchers investigated ways to avoid this issue, and strengthen the performance of the P300 BCI system.

For example, Takano et al. identified that the color of the stimuli could influence P300-speller system performance. They replaced the white/gray flicker matrix with a green/blue flicker matrix and found that the chromatic stimulus improved the performance of the P300-speller system (Takano et al., 2009). Jin et al. (2012) proposed a set of stimuli patterns that made use of images of the face with different emotional content and degrees of movement, including neutral faces, smiling faces, shaking neutral faces, and shaking smiling faces. The results revealed that BCIs that make use of face-based stimuli paradigms are superior to the traditional RCP. Kaufmann et al. (2011) attempted to overlay characters used in a P300 speller with semitransparent images of familiar faces. This resulted in a higher classification accuracy by evoking N170 and N400 ERPs. The N170 is a negative voltage deflection occurring approximately 200 ms after stimulus onset, which is generally related to motion of the stimuli (Jin et al., 2015), speech processing (Niznikiewicz and Squires, 1996), and vocabulary selection (Kutas and Hillyard, 1980). The N400 component occurs at 300-500 ms post-stimulus, and is connected with face recognition (Kaufmann et al., 2011) and language understanding (Johnson and Hamm, 2000). The influences produced by stimuli have also been reflected in other factors, such as, but not limited to, the inter-stimulus intervals (Sellers et al., 2006), stimulus intensity (Cass and Polich, 1997), and stimulus motion (Sutton et al., 1965; Martens et al., 2009). A large number of works have attempted to design optimal paradigms based on face stimuli to improve the performance of BCI systems. For example, Li et al. (2015) observed that compared with a paradigm that only used semitransparent famous faces, the green semitransparent famous face paradigm could lead to improved classification performance. Based on this, we further explore the performance differences between red semitransparent face (RSF), green semitransparent face (GSF), and blue semitransparent face (BSF) patterns. In addition, Guo et al. (2019) investigated how red, green, and blue (RGB) colors may be used as stimuli in a new layout of flash patterns based on single character presentation. They reported that the red stimuli paradigm yielded the best performance. Thus, we hypothesize that faces, that are colored red, can produce a higher classification accuracy compared to patterns that combine red, green, and blue colors with faces.

Although a large number of works have attempted to design optimal paradigms to improve the performance of BCI systems, there are scarce studies on the pattern of chromatic difference and face combination. In our new patterns, the flashing row or column in the BCI display grid is overlaid with semitransparent faces that are colored red, green, or blue and we compare the effect of these three new spelling patterns on BCI performance. In addition, we investigate the ERP waveforms induced by the proposed "red semitransparent face" (RSF), "green semitransparent face" (GSF), and "blue semitransparent face" (BSF) patterns and evaluate the classification performance among the three patterns.

MATERIALS AND METHODS

Participants

Twelve healthy participants (S1–S12, five females and seven males, aged 22–25 years, mean 24 years old) with normal or corrected to normal vision volunteered for the current study. All participants' native language is Mandarin Chinese, and they are familiar with the Western characters used in the display. They are all right-handed and had normal color vision. Before the experiment began, all participants provided informed consent via a process which the local ethics committee approved. Two participants' data was abandoned because the accuracy of three patterns were all lower than 60%. According to Kubler et al. (2004), these two participants may be described as "BCI-illiterate." Four of the ten participants (S1, S3, S6, and S7) had participated in a BCI experiment previously. All participants were informed of the whole experimental process in advance.

Experimental Design

A 20-inch LCD monitor (Lenovo LS2023WC) with standard RGB gamut and 1600 × 900 resolution was used for stimuli presentation. Its maximum luminance was set to 200cd/m^2. In the experiment, we instructed participants sit approximately 105 cm in front of the display, which was 30 cm tall (visual angle: 16.3°) and 48 cm wide (visual angle: 25.7°) in a quiet laboratory which was relatively dim with the optic intensity of the environment approximately 40lx. Participants were asked to relax themselves and avoid unnecessary movement throughout the experiment. The graphical interface of the BCI was developed

using "Qt Designer 4.8" software. The semitransparent images of faces, painted with three primary colors, red (255,0,0), green (0,255,0), and blue (0,0,255), were selected as stimuli, as shown in the **Figure 1**, and the transparency was set to 50%. The stimulus onset asynchrony (SOA) was set to 250 ms, and the stimulus interval was set to 100 ms throughout all stages of the experiment.

Figure 1A shows the interface of the 6 × 6 spelling matrix before the experiment began; it contains 26 letters and 10 digits. The parameters of the three patterns including background color, the appearance and distance of characters and the stimuli style remain the same throughout the experiment. In **Figure 1B**, the pattern showed a semitransparent face colored red as the stimulus covered the characters. For the sake of convenience, we refer to this as the RSF pattern. **Figure 1C** shows the semitransparent face colored green as the stimulus covered the characters. This is referred to as the GSF pattern. **Figure 1D** shows the semitransparent face colored blue as the stimulus covered the characters. This is called the BSF pattern. In addition, **Figures 1B–D** presented the fifth flash.

In the current study, three patterns were presented to participants in sequence. During the experiment, participants were requested to silently count the number of times target characters flashed. The stimulus presentation pattern is based on binomial coefficients (Jin et al., 2010, 2014a). The formulation is $C(n,k) = n!/k!\,(n-k)!,\ 0 \le k \le n$, where n refers to the number of flashes per trial and k refers to the number of flashes per trial for an element in the matrix. In this study, the combination of $C(12, 2)$ was used to represent the 12-flash pattern. **Table 1** describes the coding of the stimulus sequence in the 12-flash pattern with 36 flash pattern pairs. The locations in **Table 1** correspond to the locations of the 36 characters in

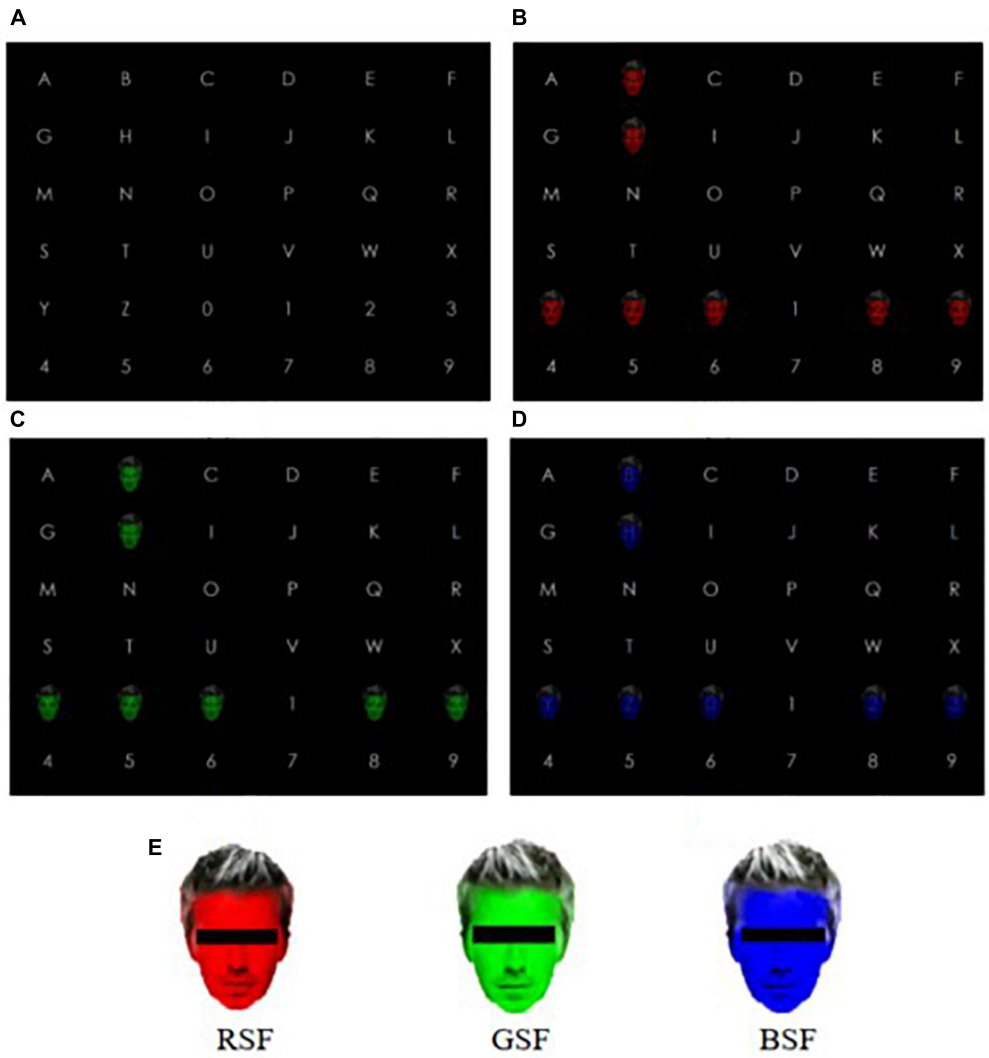


FIGURE 1 | The experimental pattern. **(A)** Character matrix; **(B)** Red semitransparent face (RSF) pattern; **(C)** Green semitransparent face (GSF) pattern; **(D)** Blue semitransparent face (BSF) pattern; **(E)** the legend of the three stimuli. Note that in order to avoid copyright infringement, faces are portrayed with censor boxes. (During the experiment censor boxes were not presented). In addition, **(B–D)** presented the fifth flash.

TABLE 1 | The coding of stimulus sequence of the 12-flash pattern.

(1,4)	(1,5)	(1,6)	(1,7)	(1,8)	(1,9)
(2,10)	(2,5)	(2,6)	(2,7)	(2,8)	(2,9)
(3,10)	(3,11)	(3,6)	(3,7)	(3,8)	(3,9)
(4,10)	(4,11)	(4,12)	(4,7)	(4,8)	(4,9)
(5,10)	(5,11)	(5,12)	(1,10)	(5,8)	(5,9)
(6,10)	(6,11)	(6,12)	(3,12)	(2,11)	(6,9)

Figure 1A. Specifically, the first pair (1,4) in **Table 1** means the first and the fourth flash will cover character "A". During the offline and online block – for each of the three patterns – the presentation sequences for each stimulus are consistent with **Table 1**.

The flow diagram of the experiment is shown in **Figure 2**. Each pattern was presented during both offline and online blocks. The offline block included three runs. In each run participants were asked to attempt to spell five targets without any break. After each offline run, participants had 3–5 min rest. Moreover, each target needed to be presented in 16 trials before it can be identified, and each trial consisted of 12 stimuli flashes. In the offline block, no feedback was displayed to the participants. The online block contained one run, which included a spelling task with 36 targets, each of which contained n trials, where n was decided by online adaptive strategy (Jin et al., 2011) for each target. Before each run began, the prompt box over the character indicated the target character.

Given the order that the three patterns were tested in could affect the performance, we kept split the participants into three, uniformly sized, groups. Each group was presented the three patterns in a different order. **Table 2** lists the order of presentation of the three patterns for all 12 participants. Specifically, participants S1, S4, S5, and S8 attempted to use the RSF pattern, followed by the GSF pattern, and then the BSF pattern. Participants S2, S3, S6, and S11 used the GSF pattern, BSF pattern, and then the RSF pattern, Finally, participants S7, S9, S10, and S12 used the BSF pattern, RSF pattern, and then the GSF pattern (see **Table 2**).

Stimulus Consistency

We prepared the interface composed of a black background and white characters, which was used to show a traditional P300 speller interface (Farwell and Donchin, 1988). In order to ensure the consistency of the color lightness and saturation across the three stimuli, we referred to G. Saravanan's study (Saravanan et al., 2016) which transformed RGB values to the Hue, Saturation, and Luminance (HLS) color scale. The conversion formula is expressed in the following equation.

$$R' = R/255; \quad G' = G/255; \quad B' = B/255 \tag{1}$$

$$C_{max} = MAX\left(R', G', B'\right) \tag{2}$$

$$C_{min} = MIN\left(R', G', B'\right) \tag{3}$$

$$\Delta = C_{max} - C_{min} \tag{4}$$

The HSL values can be calculated by the following formula.

$$H = \begin{cases} 0°, & \Delta = 0 \\ 60° \times \left(\frac{G'-B'}{\Delta} + 0\right), & C_{max} = R' \\ 60° \times \left(\frac{B'-R'}{\Delta} + 2\right), & C_{max} = G' \\ 60° \times \left(\frac{R'-G'}{\Delta} + 4\right), & C_{max} = B' \end{cases} \tag{5}$$

$$S = \begin{cases} 0, & \Delta = 0 \\ \frac{\Delta}{1-|2L-1|}, & other \end{cases} \tag{6}$$

$$L = (C_{max} + C_{min}) / 2 \tag{7}$$

In this work, we calculated the corresponding values of hue, saturation, and luminance of the three stimuli. The three stimuli refer to the red (255,0,0), green (0,255,0), and blue (0,0,255) colors. It is noteworthy that the background of the interface was black with white characters and the three stimuli were consistent in saturation and luminance while differing in hue. This is shown in **Table 3**.

Electroencephalogram Acquisition

These EEG signals were recorded with g.USBamp and g.EEGcap systems (Guger Technologies, Graz, Austria). The sample rate of the amplifier was set as 256 Hz, the sensitivity value was100μV, and a third-order Butterworth band-pass filter was applied from 0.1 to 30 Hz (Munssinger et al., 2010; Halder et al., 2016). In this paper, we chose 16 electrode positions, based on the international 10–20 system (Jin et al., 2014a), which were positioned over areas of the brain associated with vision. These electrodes were Fz, F3, F4, FC1, FC2, C3, Cz, C4, P3, Pz, P4, P7, P8, O1, Oz, and O2. The ground electrode was placed at position FPz, while the reference electrode was placed on the right mastoid (R) (Jin et al., 2010, 2012, 2015). According to Petten and Kutas (1988), the use of the right mastoid reference leads to conclusions which are somewhat similar to those with the average of left and right mastoids. The electrode impedance was kept below 5 kΩ in the experiment (Munssinger et al., 2010). **Figure 3** shows the configuration of the selected electrode positions.

Feature Extraction and Classification

After completing the offline block, feature extraction is used to reduce dimensionality and hence computation time. Extracted features were used to construct the individual classifier model, which was applied during the online block. A band pass filter was applied to filter the EEG between 1 and 30 Hz to reduce high frequency noise. The filtering algorithm we applied was a third-order Butterworth filter. In order to eliminate the impact of electrical noise, the IIR notch filter of 50 Hz was also applied. In order to decrease dimensionality of the data and complexity of the classification model, the filtered EEG data was down-sampled from 256 to 36.6 Hz by taking every 7th sample.

The first 800 ms of EEG after stimulus presentation was extracted from each channel. This resulted in a feature vector of size 16 × 29, where 16 is the number of channels we used and 29 is the number of sample points recorded on each channel after down-sampling. Moreover, we used winsorizing to remove ocular

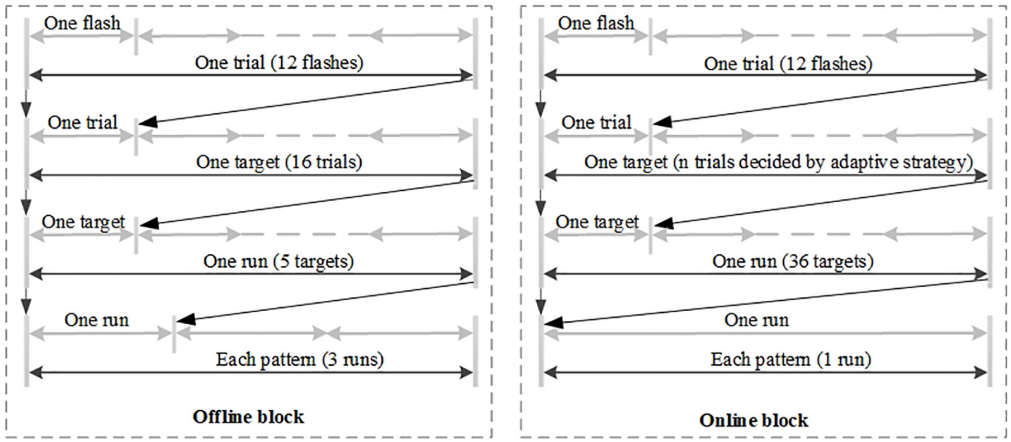


FIGURE 2 | The flow diagram of the whole experiment.

TABLE 2 | The order of patterns for 12 participants.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
RSF	1	3	3	1	1	3	2	1	2	2	3	2
GSF	2	1	1	2	2	1	3	2	3	3	1	3
BSF	3	2	2	3	3	2	1	3	1	1	2	1

Note that "1" refers to the first pattern presented to the participant, and "2" refers to the second pattern and "3" denotes the last.

TABLE 3 | The corresponding HSL value of the three stimuli.

	R (red)	G (green)	B (blue)	H (hue)	S (saturation)	L (luminance)
RSF	255	0	0	0	100	50
GSF	0	255	0	120	100	50
BSF	0	0	255	240	100	50

artifacts by filtering amplitudes which were less than or greater than 10 and 90% of the amplitude distribution across the feature set (Jin et al., 2014b).

In this study, we applied Bayesian linear discriminant analysis (BLDA) to construct the individual classifier model which was used during the online block. Due to its regularization, it can avoid the problem of overfitting of high-dimensional data or noise interference. Hoffmann et al. (2008) first proposed BLDA and applied it to the P300-based BCI system effectively. In addition, after constructing the model, the score per flash was obtained. Within one trial, that is twelve flashes, the target flash should achieve the highest mark.

In accordance with widely used standardized metrics for assessing BCI performance, the classification accuracy and information transfer rate (ITR) are applied to assess the performance of our BCI. The ITR is defined as:

$$B = log_2 N + Acc * log_2 Acc + (1 - Acc) * log_2 \frac{1 - Acc}{N - 1} \quad (8)$$

$$ITR = B * \frac{60}{T} \quad (9)$$

where N represents the total number of targets, Acc denotes the classification accuracy, and T represents the time performing each trial.

Online Adaptive System Setting

In order to improve system performance, an adaptive strategy was used with the online spelling system (Jin et al., 2011). In the online spelling system, the number of trials used to select each character is related to the classifier output after each trial. Specifically, when the classifier recognized the same character on two successive trials, no new flashes are needed and the recognized character is presented on the screen as feedback to the BCI user. If the number of trials needed to recognize a character reaches 16 without any pair of

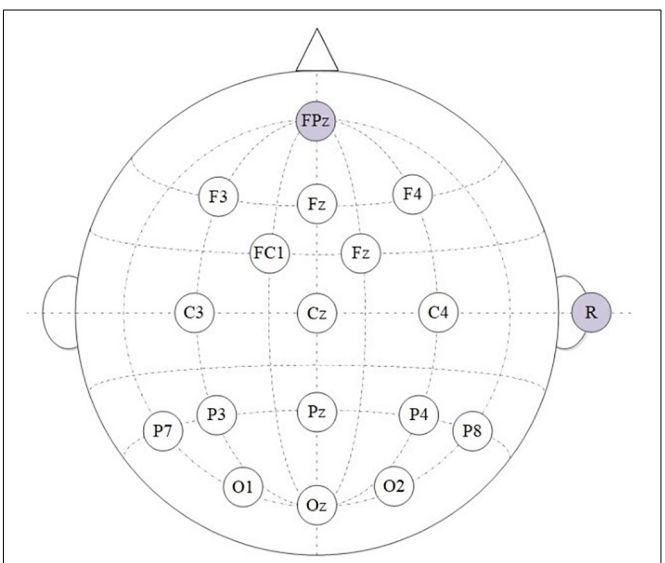


FIGURE 3 | The configuration of the selected electrode positions from the 10–20 system.

consecutive trials recognizing the same character, the classifier will automatically choose the target recognized in the final trial. For example, suppose that "A" is the target character which the classifier recognized in the first trial. If the character "A" was recognized again in the second trial, the final output will be "A". We can describe this process via $cha(n) = cha(n-1)$ $(1 < n \leq 16)$.

Statistical Analysis

The One-Sample Ryan-Joiner test based on the correction of Shapiro-Wilk was used to analyze whether the samples were normally distributed. A Repeated-measures ANOVA (RM-ANOVA) was chosen to evaluate the effect of stimuli pattern. Mauchly's test of sphericity was first used to check the data meets the assumptions of the RM-ANOVA. If the assumption was broken, Greenhouse-Geisser correction was performed to adjust the degrees of freedom. Finally, we applied Bonferroni multiple comparisons correction in *post hoc* tests (Kathner et al., 2015). The alpha level was set to 0.05 after Bonferroni correction.

RESULTS

ERP Analysis

Figure 4 illustrates the grand averaged ERP amplitudes in response to the target stimuli for ten participants over 16 electrodes across the three patterns, after applying baseline correction with a 100 ms pre-stimulus baseline. In **Figure 4**, the three colors of the curves in each channel illustrate the three different kinds of patterns respectively. Four color blocks lie around the peak point, which represents four types of potentials including the vertex positive potential (VPP), the N200, P300, and N400 potentials. We selected the latency of the potentials as the peak point with the range (min −10 ms, max + 10 ms). As we can see in **Figure 4**, the VPP components exist in frontal and central sites while the N200 and P300 components are centered over parietal and occipital areas. In addition, the peak amplitude of BSF curve performed lower than RSF and GSF curves (see **Figure 4**). According to studies of W. D. Wright (Gregory, 1973) and Fuortes (Fuortes et al., 1973), the human eye is composed of three color-sensitive cone-cell types (red, green, and blue). These three cone types have different responses for different stimulus wavelengths. Red cones are more sensitive to red color, green cones are more sensitive to green color. Among the three cone types, the red-cone presents the best response followed closely by the green-cone, with the blue cones having the lowest response, which may cause the difference. Furthermore, according to a RM-ANOVA, the P300 amplitude evoked by the RSF pattern is significantly larger than the other two patterns ($p < 0.05$) on parietal and occipital sites, corresponding to electrode P3, P7, Pz, P8, O1, Oz, and O_2.

Figure 5 shows the signed R-squared value maps from 0 to 800 ms for ten participants over 16 electrodes for each of the three patterns, which reflects the difference between the target and non-target stimuli over 16 channels. In order to show the difference among R-square map for RSF, GSF, and BSF patterns, the additional three R-square maps for the differences between RSF and GSF pattern, between RSF and BSF pattern and between GSF and BSF have also shown in **Figure 5**. The R-squared values of the ERPs evaluate the separation between target and non-target signals. The formula is given as:

$$r^2 = \left(\frac{\sqrt{N_1 N_2}}{N_1 + N_2} \cdot \frac{mean(X_1) - mean(X_2)}{std(X_1 \cup X_2)}\right)^2 \quad (10)$$

where X_1 and X_2 refer to the features of class 1 and class 2 respectively, and N_1 and N_2 are the number of corresponding samples. In **Figure 5**, the darker the color, the more distinct the features.

Classification Accuracy and Bit Rate

Figure 6 illustrates the classification accuracy and raw bit rates for each of the three patterns, which were overlapped and averaged from all trials for the ten participants based on the offline data. This valued were acquired from 15-fold cross-validation. As shown in **Figure 6**, the RSF pattern achieved the best offline accuracy and bit rate by averaging 16 trials. This pattern also used required the fewest the least trials to attain an accuracy of 100%. **Figure 7** depicts the classification accuracy based on offline single trials, which shows no significant differences across the three patterns.

In order to observe the differences between the N200, VPP, P300, and N400 ERP components between the three patterns, we chose channel P8 for measuring the N200, Cz for measuring the VPP, Pz for measuring the P300 and Cz for measuring the N400 (Farwell and Donchin, 1988; Jeffreys and Tukmachi, 1992; Duncan et al., 2009). The selected channels generally cover the highest ERP amplitude of the corresponding component.

Table 4 describes the averaged amplitudes of the VPP on channel Cz, N200 on channel P8, P300 on channel Pz and N400 on channel Cz from the peak point ± 10 ms for the ten participants. The averaged values of VPP, P300, and N400 are largest when the RSF pattern is used, and the stability of the P300 during presentation of the RSF pattern is better than that the other patterns.

Figure 8 presents the averaged contributions of the N200, P300, and N400 components to the offline classification accuracy for the ten participants. The N200 had a latency of 150–300 ms after stimulation, the P300 had a latency of 300–450 ms, and the N400 had a latency of 350–600 ms (Zhou et al., 2016). The result of the three patterns all delineated N200 and P300 played a pivotal role in offline classification. Moreover, the N400 potential has positive effect on the offline classification accuracy.

Online Analysis

Table 5 shows the online accuracies, bit rates, and the averaged numbers of trials for participants S1–S10 for each of the three patterns. The calculated *p*-values indicate the significance of the difference between each pair of accuracies. Our one-way RM-ANOVA shows a significant effect of the factor "color" on the online accuracy $\left(F(1.30, 11.65) = 8.87, p < 0.05, eta^2 = 0.50\right)$ and bit rate $(F(1.11, 10.02) = 9.25.p < 0.05, eta^2 = 0.51)$. The online accuracy of the RSF pattern was significantly higher than

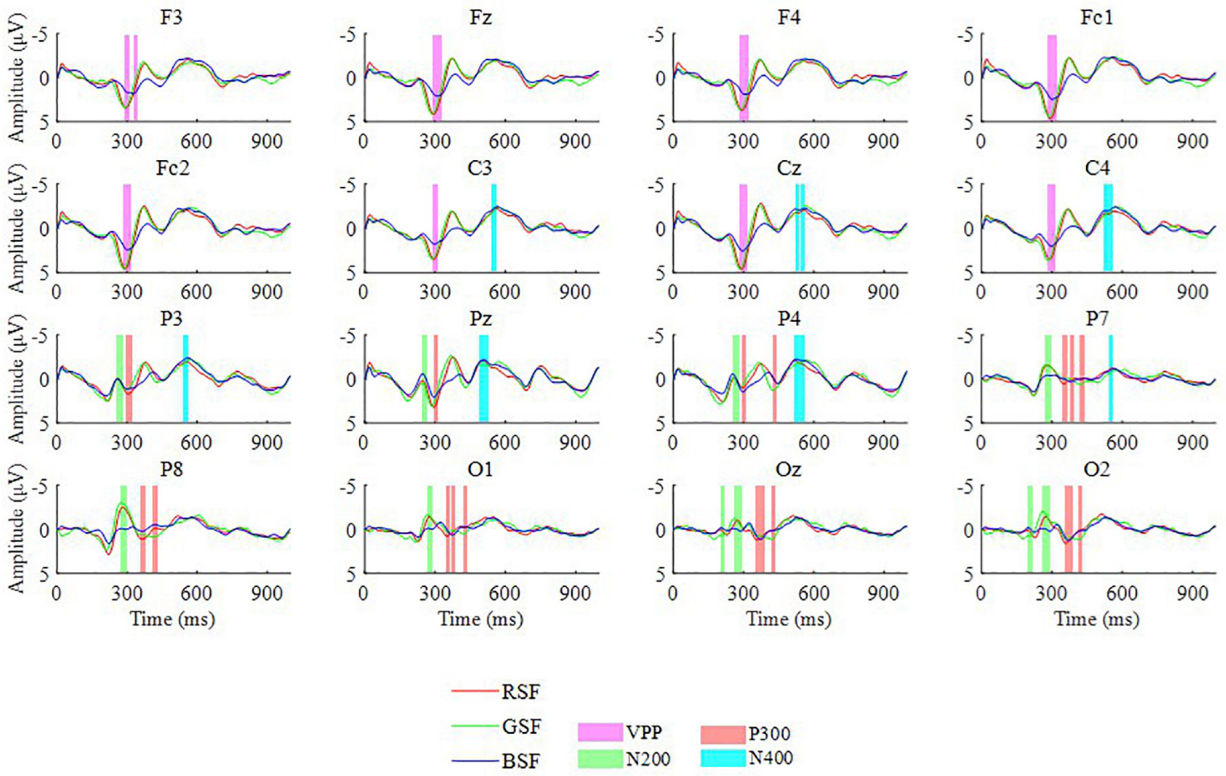


FIGURE 4 | The grand averaged ERP amplitudes of targets for 10 participants over 16 electrodes among the three patterns.

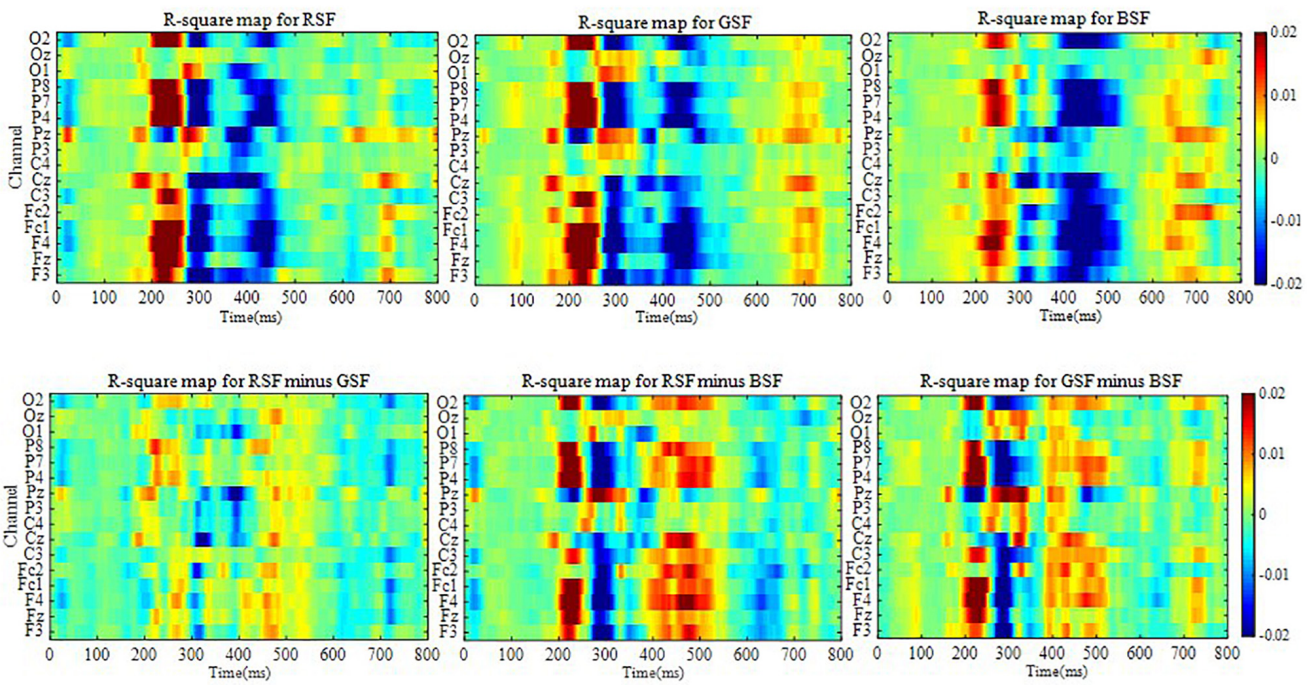


FIGURE 5 | The signed R-squared value maps from 0 to 800 ms for 10 participants over 16 electrodes for each of the three patterns and for the differences of the three patterns.

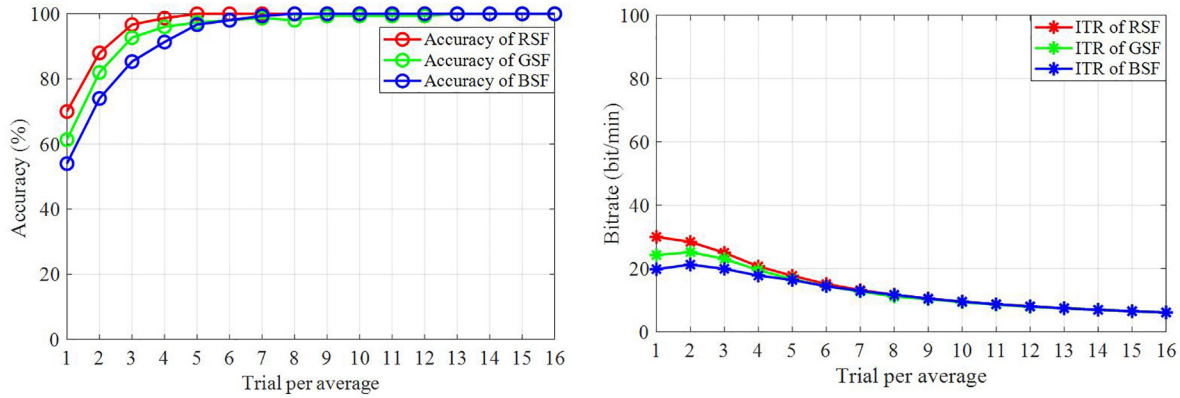


FIGURE 6 | Classification accuracy and raw bit rate based on the offline data.

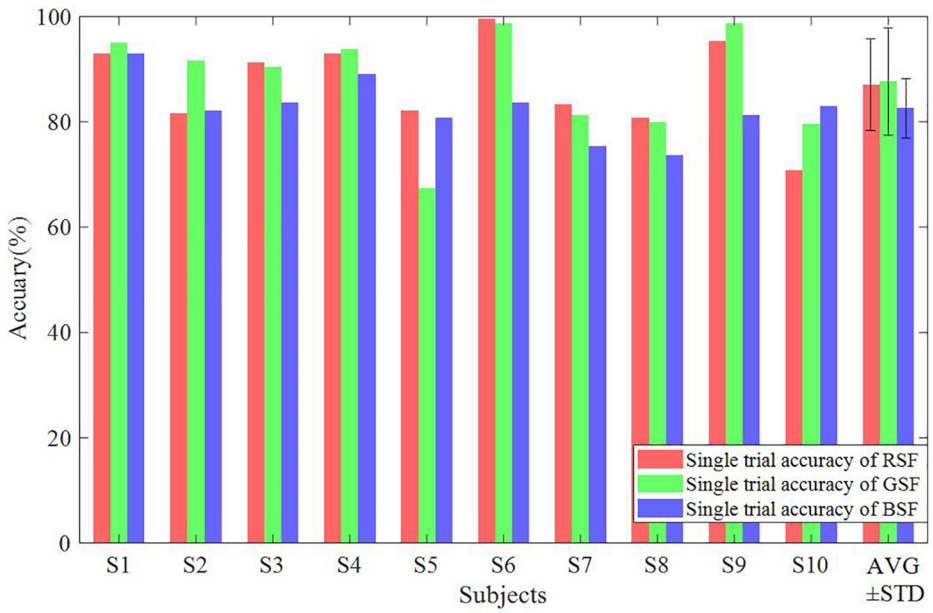


FIGURE 7 | The classification accuracy based on offline single trial classification.

that of the GSF pattern ($t = 3.24, p < 0.05, df = 9$) and the BSF pattern ($t = 4.39, p < 0.05, df = 9$). In addition, the bit rate of the RSF pattern was significantly higher than that of the GSF pattern ($t = 5.77, p < 0.05, df = 9$) and the BSF pattern ($t = 3.93, p < 0.05, df = 9$). However, there are no significant differences in the number of average trials needed for the classification across the three patterns. A boxplot of online accuracies is illustrated in **Figure 9**.

Participants' Feedback

At the end of the whole experiment, every participant was asked to grade their perception of the tiredness and difficulty of each pattern. Tiredness and difficulty were each given a rating between 1 and 3. A score of 1 corresponded to a little, a score of 2 medium, and a score of 3 quite a lot of tiredness or difficulty. The questions were asked in Mandarin Chinese. For the sake of distinguishing the differences among three patterns, a non-parametric Friedman test was applied to reveal the differences in feedback. **Table 6** delineates the feedback of all participants among three patterns.

TABLE 4 | The averaged amplitudes from each ERP peak point ± 10 ms of all participant.

		RSF (μV)		GSF (μV)		BSF (μV)	
Potential	**Channel**	**Amplitude**	**STD**	**Amplitude**	**STD**	**Amplitude**	**STD**
VPP	Cz	5.40	1.99	5.38	2.31	3.35	1.49
N200	P8	−3.67	2.10	−4.25	1.90	−1.28	1.01
P300	Pz	3.46	0.99	3.12	1.47	2.84	1.02
N400	Cz	−4.33	1.53	−4.16	1.68	−3.97	1.52

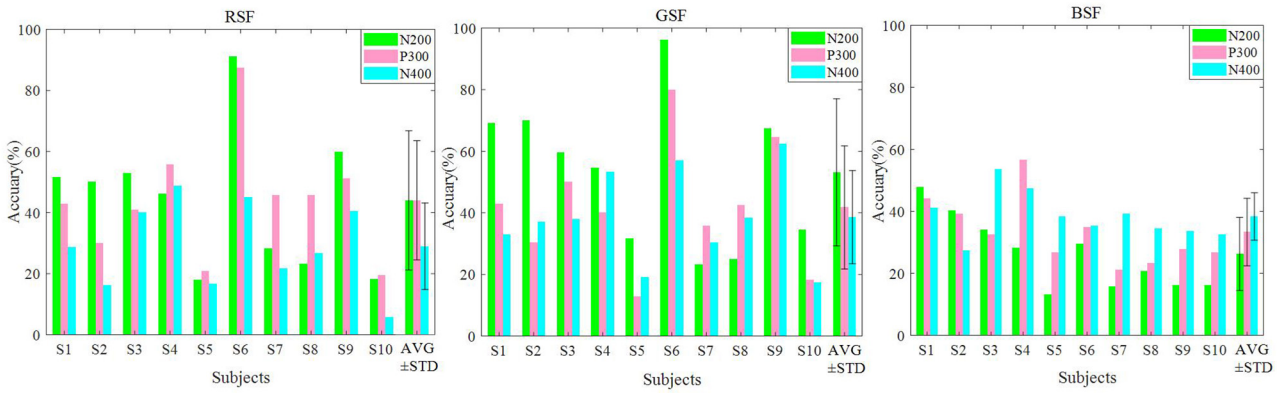


FIGURE 8 | The contributions on N200, P300, and N400 time windows on the classification accuracy.

TABLE 5 | Online accuracies, bit rates, and average trials analysis results.

	Accuracy (%)			Bit rate (bit/min)			AVT		
	RSF	**GSF**	**BSF**	**RSF**	**GSF**	**BSF**	**RSF**	**GSF**	**BSF**
S1	100.00	94.44	91.67	31.59	27.45	27.81	2.44	2.50	2.28
S2	97.22	86.11	88.89	29.60	24.66	26.32	2.44	2.31	2.28
S3	94.44	94.44	94.44	29.11	28.39	27.45	2.31	2.39	2.50
S4	100.00	94.44	91.67	33.75	31.09	28.83	2.23	2.11	2.17
S5	91.67	77.78	75.00	24.93	18.38	17.03	2.64	2.72	2.78
S6	100.00	97.22	86.11	35.74	32.26	25.40	2.06	2.17	2.22
S7	83.33	77.78	69.44	21.47	17.31	15.28	2.58	2.94	2.72
S8	88.89	88.89	72.22	24.80	22.49	16.81	2.47	2.81	2.61
S9	100.00	100.00	66.67	35.13	34.20	14.19	2.11	2.19	2.75
S10	83.33	66.67	77.78	20.63	14.64	17.93	2.72	2.64	2.81
AVG	**93.89**	87.78	81.39	**28.68**	25.09	21.70	**2.40**	2.48	2.51
STD	6.78	10.65	10.32	5.50	6.74	5.90	0.22	0.29	0.24
p	**RSF vs. GSF**	GSF vs. BSF	**RSF vs. BSF**	**RSF vs. GSF**	GSF vs. BSF	**RSF vs. BSF**	RSF vs. GSF	GSF vs. BSF	RSF vs. BSF
	0.031	0.392	**0.005**	**0.001**	0.417	**0.010**	0.535	1.000	0.464

Note that AVT refers to the average number of trials used in online spelling. The p-value is obtained by applying Bonferroni correction. The bold values refer to the highest online accuracy, bit rate, and the smallest number of trials among three patterns. In the row of "value," the bold item indicates which item is significant.

No significant difference ($\chi^2 = 5.034$, $p > 0.05$) was found between the patterns in terms of difficulty or tiredness ($\chi^2 = 0.636, p > 0.05$).

DISCUSSION

ERP-based BCI systems have been widely investigated over many years and some researchers have designed novel stimulus paradigms to optimize system performance. Previous work has indicated that familiar faces, colored green, may be used as a part of the ERP-based BCI display pattern to achieve higher performance than other display patterns, such as the familiar face pattern based in P300-speller BCI system (Li et al., 2015). Therefore, we evaluated how this paradigm was influenced by other colors (red, green, and blue).

Related studies have indicated that, when familiar faces are used as stimuli, they may strongly elicit several ERPs, including the VPP, N200, P300, and N400 components. Cheng et al. (2017) reported that the semitransparent face pattern can evoke larger N200 components, which can contribute to improving classification accuracy. Eimer (2000) revealed that familiar faces could elicit an N400 in parietal and central cortical areas. In addition, the VPP component remarkably increase for face-related stimuli over frontal and central sites (Zhang et al., 2012). Among the three patterns evaluated in this study, we found all the ERP components, shown in **Figure 4**. Moreover, we can see from **Figure 8** that the P300, N200, and the N400 all contribute to the classification accuracy. The results also indicate that the RSF pattern could elicit larger P300 potentials on parietal and occipital areas.

Generally, the performance of a BCI can be evaluated by online accuracy and ITR. The results listed in **Table 5** indicate that the RSF pattern achieved the highest online averaged accuracy of 93.89%, followed by the GSF pattern with 87.78%, while the lowest accuracy was achieved with the BSF pattern

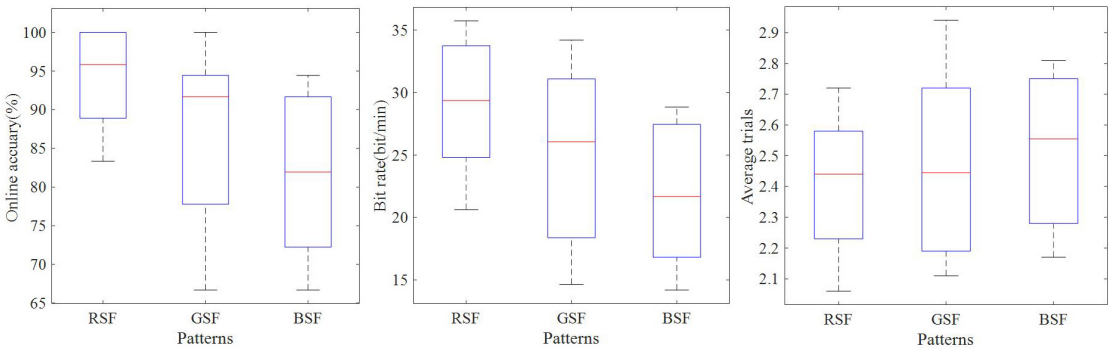


FIGURE 9 | Boxplot of online classification accuracies, bit rates, and numbers of trials used to construct the averaged ERPs.

TABLE 6 | The feedback of all participants for each of the three patterns.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	AVG ± STD
Tiredness											
RSF	2	1	1	1	2	2	3	2	1	2	1.7 ± 0.67
GSF	1	2	1	2	1	1	2	1	1	2	1.4 ± 0.52
BSF	2	2	2	1	3	2	2	3	1	3	2.1 ± 0.74
Difficulty											
RSF	1	1	1	2	2	2	2	2	1	1	1.5 ± 0.53
GSF	1	1	1	1	1	2	3	1	2	2	1.5 ± 0.71
BSF	1	2	1	1	3	2	2	2	1	2	1.7 ± 0.67

Note that a score of "1" denotes a few, "2" medium, and "3" many. AVG denotes the average and STD denotes the standard deviation.

(81.39%). Four of the participants using the RSF pattern obtained 100% online accuracy. Furthermore, the online accuracy achieved with the RSF is significantly higher than that achieved with the GSF pattern ($p < 0.05$) and the BSF pattern ($p < 0.05$). In addition, significant differences in bit rate were found between the RSF and GSF patterns ($p < 0.05$) and between RSF and BSF patterns ($p < 0.05$). The averaged bit rate of the RSF pattern was 38.45 bit/min, and the bit rate of the GSF pattern was 33.71 bit/min, while the bit rate of the BSF was 28.76 bit/min. Due to the averaged presentation order of the three patterns for all participants, the effect caused by the order of pattern presentation can be ignored. Consequently, we may conclude that the RSF pattern yielded the best performance of the three patterns.

In order to further explain the findings, it is necessary to consider relevant psychological and physiological studies. Research has shown that long-wavelength colors (e.g., red and yellow) are more arousing than short-wavelength colors (e.g., blue and green) (Wilson, 1966). In our experiment, each face stimulus was presented for more than half an hour, which may induce some effects on the emotions of the participants. Additionally, an association has been reported between colors and physiological indices of cognition. For instance, the color red is frequently associated with fire and blood which can lead to excitement and fear (Kaiser, 1984; Camgöz et al., 2004). Sorokowski and Szmajke (2011) found that red could improve performance in a target-hitting task. This result indicated that participants attempting to hit a red moving objects can achieve better performance than participants attempting to hit blue or black targets.

In previous studies, the green/blue chromatic flicker as a visual stimulus yielded an 80.6% online accuracy (Takano et al., 2009). Li et al. (2015) proposed that a translucent green familiar face spelling paradigm could achieve an 86.1% averaged online accuracy. This SSVEP-based BCI system used LEDs of four different colors (red, green, blue, and yellow) flickering at four distinct frequencies (8, 11, 13, and 15 Hz) (Mouli et al., 2013). It was observed that the red color obtained the highest accuracy and bit rate in most frequencies. Therefore, a novel spelling pattern that combines chromatic difference (RGB) with semitransparent faces resulted in consistency and efficiency in online BCI performance and offline ERP waveform detection.

CONCLUSION

In the present work, we combined chromatic difference (RGB) with semitransparent face stimuli to explore the performance of different colored stimuli patterns in an ERP based BCI system. The results demonstrated that the RSF pattern yielded the best averaged online accuracy and ITR. In future work, we will attempt to train offline models using neural networks to boost the classification performance. In addition, according to Xu's study (Xu et al., 2018), a new BCI speller based on miniature asymmetric visual evoked potentials (aVEPs) could reduce visual fatigue for users. This demonstrates the feasibility to implement an efficient BCI system. We will further explore the effect of color preference on system performance and take user-friendliness into account to improve the usability of BCI systems. This may have a clinically significant impact by increasing communication speed and accuracy of the P300-speller for patients with severe motor impairment.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of East China University of Science and Technology. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

SL was the main author to raise the idea of the manuscript, design the whole experiment, and collect the original dataset. All authors contributed to the manuscript revision, read, and approved the submitted version.

FUNDING

This work was supported by the National Key Research and Development Program 2017YFB13003002. This work was also supported in part by the Grant National Natural Science Foundation of China, under Grant Nos. 61573142, 61773164, and 91420302, the programme of Introducing Talents of Discipline to Universities (the 111 Project) under Grant B17017, and the "ShuGuang" project supported by Shanghai Municipal Education Commission and Shanghai Education Development Foundation under Grant 19SG25.

REFERENCES

Camgöz, N., Yener, C., and Güvenç, D. (2004). Effects of hue, saturation, and brightness: part 2: attention. *Color Res. Appl.* 29, 20–28. doi: 10.1002/col.10214

Cass, M., and Polich, J. (1997). P300 from a single-stimulus paradigm: auditory intensity and tone frequency effects. *Biol. Psychol.* 46, 51–65. doi: 10.1016/s0301-0511(96)05233-7

Cheng, J., Jin, J., and Wang, X. Y. (2017). Comparison of the BCI performance between the semitransparent face pattern and the traditional face pattern. *Comput. Intell. Neurosci.* 2017:1323985. doi: 10.1155/2017/1323985

Coles, M. G., and Rugg, M. D. (1995). *Event-Related Brain Potentials: An Introduction*. Oxford: Oxford University Press.

Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Naatanen, R., et al. (2009). Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clin. Neurophysiol.* 120, 1883–1908. doi: 10.1016/j.clinph.2009.07.045

Eimer, M. (2000). Event-related brain potentials distinguish processing stages involved in face perception and recognition. *Clin. Neurophysiol.* 111, 694–705. doi: 10.1016/s1388-2457(99)00285-0

Farwell, L. A., and Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr. Clin. Neurophysiol.* 70, 510–523. doi: 10.1016/0013-4694(88)90149-6

Fuortes, M. G., Schwartz, E., and Simon, E. (1973). Colour-dependence of cone responses in the turtle retina. *J. Physiol.* 234, 199–216. doi: 10.1113/jphysiol.1973.sp010341

Gregory, R. L. (1973). *Eye and Brain: The Psychology of Seeing*. New York, NY: McGraw-Hill.

Guo, M., Jin, J., Jiao, Y., Wang, X., and Cichocki, A. (2019). Investigation of visual stimulus with various colors and the layout for the oddball paradigm in ERP-based BCI. *Front. Comput. Neurosci.* 13:24. doi: 10.3389/fncom.2019.00024

Halder, S., Takano, K., Ora, H., Onishi, A., Utsumi, K., and Kansaku, K. (2016). An evaluation of training with an auditory P300 brain-computer interface for the Japanese hiragana syllabary. *Front. Neurosci.* 10:446. doi: 10.3389/fnins.2016.00446

Hoffmann, U., Vesin, J.-M., Ebrahimi, T., and Diserens, K. (2008). An efficient P300-based brain–computer interface for disabled subjects. *J. Neurosci. Methods* 167, 115–125. doi: 10.1016/j.jneumeth.2007.03.005

Jeffreys, D., and Tukmachi, E. (1992). The vertex-positive scalp potential evoked by faces and by objects. *Exp. Brain Res.* 91, 340–350.

Jin, J., Allison, B. Z., Kaufmann, T., Kubler, A., Zhang, Y., Wang, X. Y., et al. (2012). The changing face of P300 BCIs: a comparison of stimulus changes in a P300 BCI involving faces, emotion, and movement. *PLoS One* 7:e49688. doi: 10.1371/journal.pone.0049688

Jin, J., Allison, B. Z., Sellers, E. W., Brunner, C., Horki, P., Wang, X. Y., et al. (2011). An adaptive P300-based control system. *J. Neural Eng.* 8:036006. doi: 10.1088/1741-2560/8/3/036006

Jin, J., Allison, B. Z., Zhang, Y., Wang, X. Y., and Cichocki, A. (2014a). An ERP-based BCI using an oddball paradigm with different faces and reduced errors in critical functions. *Int. J. Neural Syst.* 24:1450027. doi: 10.1142/S0129065714500270

Jin, J., Daly, I., Zhang, Y., Wang, X. Y., and Cichocki, A. (2014b). An optimized ERP brain-computer interface based on facial expression changes. *J. Neural Eng.* 11:036004. doi: 10.1088/1741-2560/11/3/036004

Jin, J., Horki, P., Brunner, C., Wang, X. Y., Neuper, C., and Pfurtscheller, G. (2010). A new P300 stimulus presentation pattern for EEG-based spelling systems. *Biomed. Tech.* 55, 203–210. doi: 10.1515/BMT.2010.029

Jin, J., Sellers, E. W., Zhou, S. J., Zhang, Y., Wang, X. Y., and Cichocki, A. (2015). A P300 brain-computer interface based on a modification of the mismatch negativity paradigm. *Int. J. Neural Syst.* 25:1550011. doi: 10.1142/S0129065715500112

Johnson, B. W., and Hamm, J. P. (2000). High-density mapping in an N400 paradigm: evidence for bilateral temporal lobe generators. *Clin. Neurophysiol.* 111, 532–545. doi: 10.1016/s1388-2457(99)00270-9

Kaiser, P. K. (1984). Physiological response to color: a critical review. *Color Res. Appl.* 9, 29–36. doi: 10.1002/col.5080090106

Kathner, I., Kubler, A., and Halder, S. (2015). Rapid P300 brain-computer interface communication with a head-mounted display. *Front. Neurosci.* 9:207. doi: 10.3389/fnins.2015.00207

Kaufmann, T., Schulz, S. M., Grunzinger, C., and Kubler, A. (2011). Flashing characters with famous faces improves ERP-based brain-computer interface performance. *J. Neural Eng.* 8:056016. doi: 10.1088/1741-2560/8/5/056016

Kubler, A., Neumann, N., Wilhelm, B., Hinterberger, T., and Birbaumer, N. (2004). Brain-computer predictability of brain-computer communication. *J. Psychophysiol.* 18, 121–129.

Kutas, M., and Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–205. doi: 10.1126/science.7350657

Li, Q., Liu, S., Li, J., and Bai, O. (2015). Use of a green familiar faces paradigm improves P300-speller brain-computer interface performance. *PLoS One* 10:e0130325. doi: 10.1371/journal.pone.0130325

Martens, S. M. M., Hill, N. J., Farquhar, J., and Scholkopf, B. (2009). Overlap and refractory effects in a brain-computer interface speller based on the visual P300 event-related potential. *J. Neural Eng.* 6:026003. doi: 10.1088/1741-2560/6/2/026003

Mouli, S., Palaniappan, R., Sillitoe, I. P., and Gan, J. Q. (2013). "Performance analysis of multi-frequency SSVEP-BCI using clear and frosted colour LED stimuli," in *Proceedings of the 13th IEEE International Conference on Bioinformatics and Bioengineering (Bibe)*, Chania.

Munssinger, J. I., Halder, S., Kleih, S. C., Furdea, A., Raco, V., Hosle, A., et al. (2010). Brain painting: first evaluation of a new brain-computer interface application with ALS-patients and healthy volunteers. *Front. Neurosci.* 4:182. doi: 10.3389/fnins.2010.00182

Niznikiewicz, M., and Squires, N. K. (1996). Phonological processing and the role of strategy in silent reading: behavioral and electrophysiological evidence. *Brain Lang.* 52, 342–364. doi: 10.1006/brln.1996.00016

Petten, C. V., and Kutas, M. (1988). The use of event-related potentials in the study of brain asymmetries. *Int. J. Neurosci.* 39, 91–99. doi: 10.3109/00207458808985695

Saravanan, G., Yamuna, G., and Nandhini, S. (2016). "Real Time implementation of RGB to HSV/HSI/HSL and its reverse color space models," in *Proceedings of the International Conference on Communication and Signal*, Vol. 1 (Jakarta: ICCSP), 462–466.

Sellers, E. W., Krusienski, D. J., McFarland, D. J., Vaughan, T. M., and Wolpaw, J. R. (2006). A P300 event-related potential brain-computer interface (BCI): the effects of matrix size and inter stimulus interval on performance. *Biol. Psychol.* 73, 242–252. doi: 10.1016/j.biopsycho.2006.04.007

Sorokowski, P., and Szmajke, A. (2011). The influence of the "Red Win" effect in sports: a hypothesis of erroneous perception of opponents dressed in red-Preliminary test. *Hum. Mov.* 12, 367–373.

Sutton, S., Braren, M., Zubin, J., and John, E. (1965). Evoked-potential correlates of stimulus uncertainty. *Science* 150, 1187–1188. doi: 10.1126/science.150.3700.1187

Takano, K., Komatsu, T., Hata, N., Nakajima, Y., and Kansaku, K. (2009). Visual stimuli for the P300 brain-computer interface: a comparison of white/gray and green/blue flicker matrices. *Clin. Neurophysiol.* 120, 1562–1566. doi: 10.1016/j.clinph.2009.06.002

Van Dinteren, R., Arns, M., Jongsma, M. L. A., and Kessels, R. P. C. (2014). Combined frontal and parietal P300 amplitudes indicate compensated cognitive processing across the lifespan. *Front. Aging Neurosci.* 6:294. doi: 10.3389/fnagi.2014.00294

Vidal, J. J. (1973). Toward direct brain-computer communication. *Annu. Rev. Biophys. Bioeng.* 2, 157–180. doi: 10.1146/annurev.bb.02.060173.001105

Vidal, J. J. (1977). Real-time detection of brain events in EEG. *Proc. IEEE* 65, 633–641. doi: 10.1109/proc.1977.10542

Wilson, G. D. (1966). Arousal properties of red versus green. *Percept. Mot. Skills* 23(3 Pt 1), 947–949. doi: 10.2466/pms.1966.23.3.947

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 767–791.

Xu, M. P., Xiao, X. L., Wang, Y. J., Qi, H. Z., Jung, T. P., and Ming, D. (2018). A brain-computer interface based on miniature-event-related potentials induced by very small lateral visual stimuli. *IEEE Trans. Biomed. Eng.* 65, 1166–1175. doi: 10.1109/Tbme.2018.2799661

Zhang, Y., Zhao, Q. B., Jin, J., Wang, X. Y., and Cichocki, A. (2012). A novel BCI based on ERP components sensitive to configural processing of human faces. *J. Neural Eng.* 9:026018. doi: 10.1088/1741-2560/9/2/026018

Zhou, S. J., Jin, J., Daly, I., Wang, X. Y., and Cichocki, A. (2016). Optimizing the face paradigm of BCI system by modified mismatch negative paradigm. *Front. Neurosci.* 10:444. doi: 10.3389/fnins.2016.00444

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 23 June 2020
doi: 10.3389/fnins.2020.00627

BETA: A Large Benchmark Database Toward SSVEP-BCI Application

Bingchuan Liu[1†], *Xiaoshan Huang*[1†], *Yijun Wang*[2], *Xiaogang Chen*[3] *and Xiaorong Gao*[1*]

[1] *Department of Biomedical Engineering, Tsinghua University, Beijing, China,* [2] *State Key Laboratory on Integrated Optoelectronics, Institute of Semiconductors, Chinese Academy of Sciences, Beijing, China,* [3] *Institute of Biomedical Engineering, Chinese Academy of Medical Sciences and Peking Union Medical College, Tianjin, China*

The brain-computer interface (BCI) provides an alternative means to communicate and it has sparked growing interest in the past two decades. Specifically, for Steady-State Visual Evoked Potential (SSVEP) based BCI, marked improvement has been made in the frequency recognition method and data sharing. However, the number of pubic databases is still limited in this field. Therefore, we present a **BE**nchmark database **T**owards BCI **A**pplication (BETA) in the study. The BETA database is composed of 64-channel Electroencephalogram (EEG) data of 70 subjects performing a 40-target cued-spelling task. The design and the acquisition of the BETA are in pursuit of meeting the demand from real-world applications and it can be used as a test-bed for these scenarios. We validate the database by a series of analyses and conduct the classification analysis of eleven frequency recognition methods on BETA. We recommend using the metric of wide-band signal-to-noise ratio (SNR) and BCI quotient to characterize the SSVEP at the single-trial and population levels, respectively. The BETA database can be downloaded from the following link http://bci.med.tsinghua.edu.cn/download.html.

Keywords: brain-computer interface (BCI), steady-state visual evoked potential (SSVEP), electroencephalogram (EEG), public database, frequency recognition, classification algorithms, signal-to-noise ratio (SNR)

OPEN ACCESS

Edited by:
Ian Daly,
University of Essex, United Kingdom

Reviewed by:
Yufeng Ke,
Tianjin University, China
Victor Manuel Pulgar,
Wake Forest School of Medicine, United States

****Correspondence:***
Xiaorong Gao
gxr-dea@tsinghua.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Neural Technology, a section of the journal Frontiers in Neuroscience

Received: *21 March 2020*
Accepted: *20 May 2020*
Published: *23 June 2020*

Citation:
Liu B, Huang X, Wang Y, Chen X and Gao X (2020) BETA: A Large Benchmark Database Toward SSVEP-BCI Application. Front. Neurosci. 14:627. doi: 10.3389/fnins.2020.00627

1. INTRODUCTION

The brain-computer interface (BCI) provides a new way for brain interaction with the outside world, and it is based on measuring and converting brain signals to the external commands without involving the peripheral nervous system (Wolpaw et al., 2002). The BCI technology has considerable scientific significance and application prospects, especially in the rehabilitation field (Ang and Guan, 2013; Lebedev and Nicolelis, 2017) and as an alternative access method for physically disabled people (Gao et al., 2003; Pandarinath et al., 2017). The Steady-State Visual Evoked Potential (SSVEP) represents a stable neural response elicited by periodic visual stimuli, and its frequency tagging attribute can be leveraged in the BCI (Cheng et al., 2002; Norcia et al., 2015). Among a variety of BCI paradigms, the SSVEP-based BCI (SSVEP-BCI) has gained widespread attention due to its characteristics of non-invasiveness and high signal-to-noise ratio (SNR) and information transfer rate (ITR) (Bin et al., 2009; Chen et al., 2015a). Generally, the high-speed performance of the BCI is accomplished by a multi-target visual speller, which achieves a reportedly average online ITR of 5.42 bit per second (bps) (Nakanishi et al., 2018). Besides, the ease of use and significantly lower rate of the BCI illiteracy (Lee et al., 2019) make it a promising candidate for real-world applications.

In order to improve the performance of the BCI, rapid progress has been made to facilitate frequency recognition of the SSVEP (Zerafa et al., 2018). Based on whether a calibration or training

phase is required for the extraction of spatial filters, the signal detection methods can be categorized into supervised methods and training-free methods. The supervised methods exploit an optimal spatial filter by a training procedure and achieve the state-of-the-art classification performance in the SSVEP-based BCI (Nakanishi et al., 2018; Wong et al., 2020a). These spatial filters or projection direction can be learned by exploiting individual training template (Bin et al., 2011), reference signal optimization (Zhang et al., 2013), inter-frequency variation (Yin et al., 2015), and ensemble reference signals (Nakanishi et al., 2014; Chen et al., 2015a) in the framework of canonical correlation analysis (CCA). Recently, the task-related components (Nakanishi et al., 2018) and the multiple neighboring stimuli (Wong et al., 2020a) have been utilized to derive spatial filters in order to boost the discriminative power of the learned model further. On the other hand, the training-free methods perform feature extraction and classification in one step without the training session in the online BCI. This line of work usually use a sinusoidal reference signal, and the detection statistics can be derived from the canonical correlation (Bin et al., 2009) and its filter-bank form (Chen et al., 2015b), noise energy minimization (Friman et al., 2007), synchronization index maximization (Zhang et al., 2014), and additional spectral noise estimation (Abu-Alqumsan and Peer, 2016).

Along with the rapid development of frequency recognition methods, continuous efforts have been devoted to share the SSVEP database (Bakardjian et al., 2010; Kolodziej et al., 2015; Kalunga et al., 2016; Kwak et al., 2017; Işcan and Nikulin, 2018) and contribute to public SSVEP database (Wang et al., 2017; Choi et al., 2019; Lee et al., 2019). Wang et al. (2017) benchmarked a 40-target database comprising 64-channel 5-s SSVEP trials of 35 subjects who performed the offline cue-spelling task in six blocks. Recently, Lee et al. (2019) have released a larger database of 54 subjects performing the 4-target offline and online task, and 62-channel 4-s SSVEP data were obtained having 50 trials per class. Choi et al. (2019) also provided a 4-target database, including physiological data and the 6-s SSVEP data which are collected from 30 subjects at three different frequency bands (low: 1–12 Hz; middle: 12–30 Hz; high: 30–60 Hz) during 2 days. Nevertheless, the number of public databases in the SSVEP-BCI community is still limited compared to other domains, such as computer vision, where a growing number of databases plays a critical role in the development of the discipline (Russakovsky et al., 2015). Compared to the other BCI paradigms, e.g., the motor imagery BCI, the SSVEP-BCI databases are also scarce (Choi et al., 2019). Therefore, more databases are need in the SSVEP-BCI field for the design and evaluation of methods.

To this end, we present a large **BE**nchmark database **T**owards SSVEP-BCI **A**pplication (BETA) in this study. The BETA database includes the data of 70 subjects performing the cued-spelling task. As an extension of the benchmark database (Wang et al., 2017), the number of targets is 40, and the frequency range is from 8 to 15.8 Hz. A key feature of the proposed BETA database is that it is developed for real-world applications. Different from the benchmark database, the BETA consists of the data collected outside the laboratory setting of the electromagnetic shielding room. Since it is imperative to reduce the calibration time from a practical perspective, the number of blocks is set to four instead of six that are used in the benchmark. A QWERT virtual keyboard is presented in flickers to approximate the conventional input device better and enhance user experience. To the best of our knowledge, so far, the BETA database has the largest number of subjects for the SSVEP-BCI. Since a larger database can capture the inter-subject variability better, the BETA database makes it possible to reflect a more realistic EEG distribution and potentially meet the demands of real-world BCI applications.

The remaining of the paper is organized as follows. First, the data acquisition and curation procedures are presented in section 2. The data record and availability are described in section 3. In section 4, data validation is performed, and 11 frequency recognition methods are compared on BETA. We discuss additional findings from the database in section 5. Finally, the conclusions are given in section 6.

2. MATERIALS AND METHODS

2.1. Participants

Seventy healthy volunteers (42 males and 28 females) with an average age of 25.14 ± 7.97 (mean ± standard deviation, ranging from 9 to 64 years) participated in our study. All the participants had a normal or corrected to normal vision, and they all signed a written consent before the experiment; for the participants under 16 years old, the consent was signed by their parents. The study was carried out in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of Tsinghua University (No. 20190002).

2.2. Recruitment and Inclusion Criteria

Participants were recruited on a national scale to take part in the Brain-Computer Interface 2018 Olympics in China. The competition was held to contest and award individuals with a high performance of the BCI (SSVEP, P300, and Motor Imagery). The 70 participants who participated in this study have also participated in the second round of the contest (SSVEP-BCI track), and none of them was naive to the SSVEP-BCI. Before the enrollment, participants were informed that the data would be used in non-commercial scientific research. Participants who conformed to the experimental rules in the first round and were available for the second round planed by the contest schedule were included in the second round. All the participants met the following criteria: (1) they had no history of epileptic seizures or other neuropsychiatric disorders, (2) they had no attention-deficit or hyperactivity disorder, and (3) they had no history of brain injury or intracranial implantation.

2.3. Visual Speller

This study designed a 40-target BCI speller for visual stimulation. In order to improve user experience, a graphical interface was designed to resemble the traditional QWERT keyboard. The virtual keyboard was presented on a 27-inch LED monitor (ASUS MG279Q Gaming Monitor, 1,920 × 1,080 pixels) with a refresh rate of 60 Hz. As illustrated in **Figure 1A**, 40 targets, including 10 numbers, 26 alphabets, and 4 non-alphanumeric signs (dot, comma, backspace < and space _) were aligned in five rows,

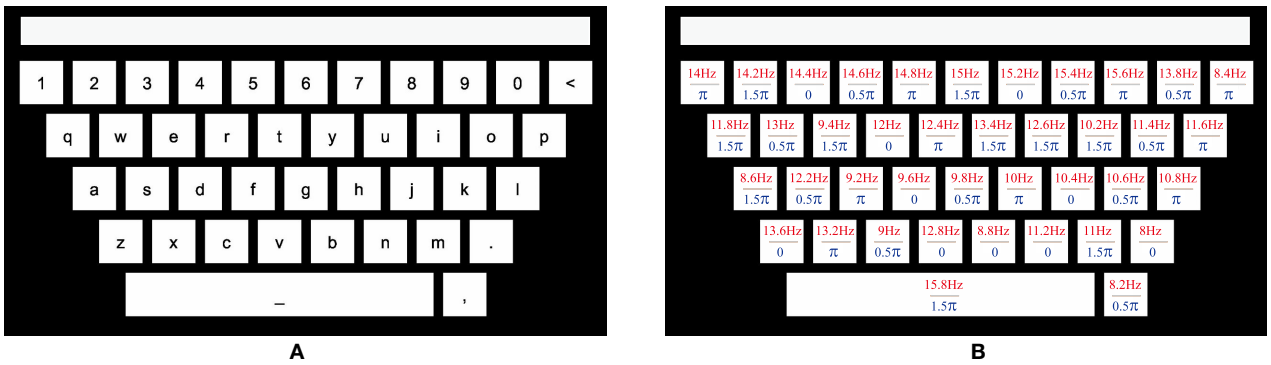


FIGURE 1 | The QWERT virtual keyboard for a 40-target BCI speller. **(A)** The layout of a conventional keyboard with ten numbers, 26 alphabets and four non-alphanumeric keys (dot, comma, backspace <, and space _) aligned in five rows. The upper rectangle is designed to present the input character. **(B)** The frequency and initial phase of each target are encoded using the joint frequency and phase modulation.

with a spacing of 30 pixels. The stimuli had the dimension of 136 × 136 pixels (3.1°× 3.1°) for the square, and 966 × 136 pixels (21°× 3.1°) for the space rectangle. The topmost blank rectangle was for result feedback (**Figure 1A**).

A sampled sinusoidal stimulation method (Manyakov et al., 2013; Chen et al., 2014) was adopted to present the visual flicker on the screen. In general, the stimulus sequence of each flicker can be generated by

$$s(f, \phi, i) = \frac{1}{2}\{1 + \sin[2\pi f(i/RefreshRate) + \phi]\} \quad (1)$$

where i denotes the frame index in the stimulus sequence, and f and ϕ denote the frequency and phase values of the encoded flicker that uses a joint frequency and phase modulation (JFPM) (Chen et al., 2015a). The grayscale value of the stimulus sequence ranges from 0 to 1, where 0 indicates dark, and 1 indicates the highest luminance of the screen. For the 40 targets, the tagged frequency and phase values can be respectively obtained by

$$\begin{aligned} f_k &= f_0 + (k-1)\cdot\Delta f \\ \Phi_k &= \Phi_0 + (k-1)\cdot\Delta\Phi \end{aligned} \quad (2)$$

where the frequency interval Δf is 0.2 Hz, the phase interval $\Delta\Phi$ is 0.5 π, and k denotes the index from dot, comma, and backspace, followed by a to z and 0–9, and space. In this work, f_0 and Φ_0 are set to 8 Hz and 0, respectively. The parameters of each target are presented in **Figure 1B**. The stimulus was presented by MATLAB (MathWorks, Inc.) using Psychophysics Toolbox Version 3 (Brainard, 1997).

2.4. Procedure

This study includes four blocks of online BCI experiments with a cued-spelling task. The experiments were as follows. Each block consisted of 40 trials, and there was one trial for each stimulus target in a randomized order. Trials began with a 0.5 s cue (a red square covering the target) for gaze shift, which was followed by flickering on all the targets, and ended with a rest time of 0.5 s. The participants were asked to avoid eye blinking during the flickering process. During the 0.5 s rest, the resulting feedback, which represented one of the recognized characters, was presented in the topmost rectangle after online processing by a modified version of the FBCCA method (Chen et al., 2015b). For the first 15 participants (S1–S15), the flickering lasted at least 2 s, and for the remaining 55 participants (S16–S70), it lasted at least 3 s. In order to avoid visual fatigue, there was a short break between two consecutive blocks.

2.5. Data Acquisition

The 64-channel EEG data were recorded by SynAmps2 (Neuroscan Inc.) according to the international 10-10 system. The sampling rate was set 1,000 Hz, and the pass-band of the hardware filter was 0.15–200 Hz. A built-in notch filter was applied to remove the 50 Hz power-line noise. The event triggers were sent from the stimulus computer to the EEG amplifier and synchronized to the EEG data by a parallel port as an event channel. The impedance of all the electrodes was kept below 10 kΩ. The vertex electrode Cz was used as a reference. During the online experiment, nine parietal and occipital channels (Pz, PO3, PO5, PO4, PO6, POz, O1, Oz, and O2) were selected for online analysis to provide the feedback result. In order to record the EEG data in real-world scenarios, the data were recorded outside the electromagnetic shielding room.

2.6. Data Preprocessing

According to the previous study (Chen et al., 2015a,b), the SSVEP harmonics in this paradigm have a frequency range of up to around 90 Hz. Based on the finding, a band-pass filtering (i.e., zero-phase forward and reverse filtering using eegfilt in EEGLAB (Delorme and Makeig, 2004) between 3 and 100 Hz was conducted to remove the environmental noise. Then, the epochs were extracted from each block, and they included 0.5 s before the stimulus onset, 2 s (for S1–S15) or 3 s (for S16–S70) of the stimulation, and 0.5 s after the simulation. The last 0.5 s of the epochs could contain the SSVEP data if the duration of the trial was > 2 s (for S1–S15) or 3 s (for S16–S70). Since frequency resolution could not affect the classification result of the SSVEP

(Nakanishi et al., 2017), all the epochs were then down-sampled to 250 Hz.

2.7. Metrics

The SSVEP data quality was evaluated quantitatively by the signal-to-noise ratio (SNR) analysis and classification analysis. As for the SNR-based analysis, in most of the previous studies (Chen et al., 2015a,b; Xing et al., 2018), the narrow-band SNR metric was used. The narrow-band SNR (in decibels, dB) can be defined as a ratio of the spectral amplitude at the stimulus frequency to the mean value of the ten neighboring frequencies (Chen et al., 2015b)

$$SNR = 20\log_{10}\frac{y(f)}{\sum_{k=1}^{5}[y(f-\Delta f\cdot k)+y(f+\Delta f\cdot k)]} \quad (3)$$

where $y(f)$ denotes the amplitude spectrum at frequency f calculated by the Fast Fourier Transform (FFT), and Δf denotes the frequency resolution.

Along with the narrow-band SNR, we used the wide-band SNR as a primary metric to characterize better both the wide-band noise and the contribution of harmonics to the signals. The wide-band SNR (in decibels, dB) can be defined as:

$$SNR = 10\log_{10}\frac{\sum_{k=1}^{k=Nh}P(k\cdot f)}{\sum_{f=0}^{f=f_s/2}P(f)-\sum_{k=1}^{k=Nh}P(k\cdot f)} \quad (4)$$

where Nh denotes the number of harmonics, $P(f_n)$ denotes the power spectrum at frequency f, and $f_s/2$ represents the Nyquist frequency. In the wide-band SNR, the sum of power spectrum of multiple harmonics ($Nh = 5$) is regarded as the signal and the energy of full spectral band subtracted from the signal is considered as noise.

The classification accuracy and the information transfer rate (ITR) have been widely used in the BCI community to evaluate the performance of different subjects and algorithms. The ITR (in bits per min—bpm) can be obtained by (Wolpaw et al., 2002):

$$ITR = 60\cdot(\log_2 M + P\log_2 P + (1-P)\log_2\frac{1-P}{M-1})/T \quad (5)$$

where M denotes the number of classes, P denotes the classification accuracy, and T (in seconds) denotes the average target selection time. The variable T in the equation represents the sum of gaze time and overall gaze shift time. To calculate the theoretical ITR for offline analysis, a gaze shift time of 0.55 s is chosen according to the previous studies (Chen et al., 2015b; Wang et al., 2017), which was proven sufficient in an online spelling task (Chen et al., 2015b).

2.8. Statistical Analysis

A linear regression was conducted to understand the relationship between the SNR and ITR metrics. To meet the assumptions of linear regression, the following procedures were conducted. A scatter plot of SNR against ITR was diagrammed to establish the linearity by visual inspection. The independence of residuals was ascertained by using the Durbin-Watson test. The standardized residuals were checked in the range of ±3 to ensure that there were no outliers in the data. The homoscedascity was ensured by assessing a plot of standardized residuals versus standardized predicted values. The normality of residuals was guaranteed by assessing a normal probability plot. The R^2 and adjusted R^2 were calculated to reflect the goodness-of-fit of the regression model. The statistical significance of the model is evaluated by analysis of variance (ANOVA).

The ITR values obtained from different methods were compared using a one-way repeated-measures ANOVA with a within-subject factor of method. A Greenhouse-Geisser correction was applied if the sphericity was violated, as assessed by Mauchly's test of sphericity. When there was a significant main effect ($p < 0.05$), *post-hoc* paired-sample t-tests were performed and Bonferroni adjustment was applied for multiple comparison. To reflect the effect size, partial eta-squared (η^2) was calculated. A Mann-Whitney U test was conducted to determine if there were differences in the SNR metrics. All the statistical procedures were processed using SPSS Statistics 20 (IBM, Armonk, NY, USA). Data were presented as mean ± standard error of the mean (s.e.m.) unless otherwise stated.

3. RECORD DESCRIPTION

The database used in this work is freely available for scientific research, where it is stored in the MATLAB .mat format. This database contains 70 subjects, and each subject corresponds to one mat file. The names of subjects are mapped to indices from S1 to S70 for de-identification. Each file in the database consists of a MATLAB structure array, which included the 4-block EEG data and its counterpart supplementary information as its fields. The website for accessing the database is http://bci.med.tsinghua.edu.cn/download.html.

3.1. EEG Data

After data preprocessing, the EEG data were store as a 4-way tensor, with a dimension of channel × time point × block × condition. Each trial included the 0.5-s data before the event onset, and the 0.5-s data after the time window of 2 or 3 s. For S1–S15, the time window was 2 s, and the trial length was 3 s, whereas for S16–S70, the time window was 3 s and the trial length was 4 s. Additional information about the channel and condition can be found in the following section about the supplementary information.

3.2. Supplementary Information

The supplementary information is comprised of personal information, channel information, BCI quotient, SNR, sampling rate, and each condition's frequency and initial phase. The personal information contained information about the age and gender of a subject. The channel information denoted a location matrix (64 × 4), where the first column represented the channel index, and the second and third columns represented the degree and radius in polar coordinates, respectively; and the last column represented the channel name. The SNR information consisted of the mean narrow-band SNR and wide-band SNR values of

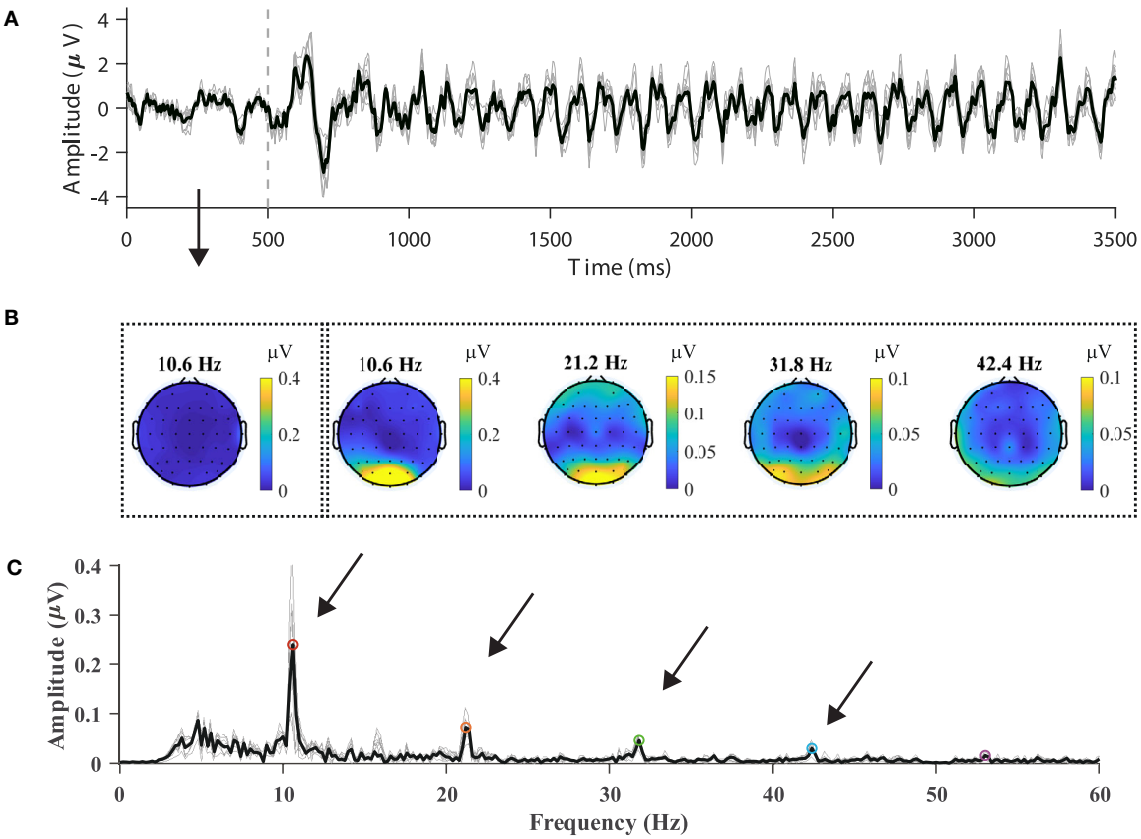


FIGURE 2 | Typical SSVEP features in the temporal, spectral, and spatial domains. **(A)** Time course of average 10.6-Hz SSVEP of nine parietal and occipital channels (Pz, PO3, PO5, PO4, PO6, POz, O1, Oz, and O2). The dash line represents stimulus onset. **(B)** The topographic maps of SSVEP amplitudes at frequencies in the range from the fundamental signal (10.6 Hz) to the fourth harmonic (21.2, 31.8, and 42.4 Hz). The leftmost scalp map indicates the spectral amplitude at the fundamental frequency before stimulus. **(C)** The amplitude spectrum of the SSVEP of the nine channels at 10.6 Hz. Up to five harmonics are visible in the amplitude spectrum. The averaged spectrum across channels is represented in the dark line in **(A,C)**.

each subject, which were calculated by Equations (3) and (4), respectively. The initial phase was given in radius.

4. DATA EVALUATION

4.1. Temporal, Spectral, and Spatial Analysis

In order to validate the data quality by visual inspection, nine parietal and occipital channels (Pz, PO3, PO5, PO4, PO6, POz, O1, Oz, and O2) were selected, and epochs were averaged with respect to the channels, blocks, and subjects. For the sake of consistency regarding the data format, the subjects from S16 to S70 were chosen for analysis. **Figure 2A** illustrates the averaged temporal amplitude at the stimulus frequency of 10.6 Hz. After a delay, which was in the range of 100–200 ms, at the stimulus onset, a steady-state and time-locked characteristic could be observed in the temporal sequence, as shown in **Figure 2A**. The data between 500 and 3,500 ms were extracted and padded with 2,000 ms zeros, yielding a 0.2 Hz spectral resolution, as shown in **Figure 2C**. In the amplitude spectrum, the fundamental frequency (10.6 Hz: 0.266 μV) and three harmonics (21.2 Hz: 0.077 μV, 31.8 Hz: 0.054 μV, 42.4 Hz: 0.033 μV) could be distinguishable from the background EEG. Note that at high frequencies ($>$ 60 Hz), the amplitude of both the harmonic signals and noise was small due to the volume conduct effect (van den Broek et al., 1998), which is why they are not shown in **Figure 2C**.

Figure 2B illustrates the topographic mappings of the spectrum at frequencies in the range from the fundamental signal to the fourth harmonic. The result presented in **Figure 2B** indicates that fundamental and harmonic signals of the SSVEP are distributed predominantly in the parietal and occipital regions. The frontal and temporal regions of the topographic maps also show an increase in the spectrum, which can represent noise or SSVEP oscillation from the occipital region (Thorpe et al., 2007; Liu et al., 2017). In order to characterize the response property of the SSVEP, the amplitude spectrum is represented as a function of stimulus frequency in **Figure 3**. According to the amplitude spectrum, the spectral response of the SSVEP decreased rapidly with the number of harmonics; namely, up to five harmonics are visible. A dark line at the response frequency of 50 Hz results from the notch filtering. A bright line at the

15.8 Hz response frequency can be distractor stimulus from the SPACE target with a larger size.

4.2. SNR Analysis

As a metric independent of different classification algorithms, the SNR measures available stimulus-evoked components in the SSVEP spectrum. In the SNR-based analysis, the BETA database was compared with the benchmark database of the SSVEP-based BCI (Wang et al., 2017). The narrow-band and wide-band SNR values were calculated for each trial by Equations (3) and (4), respectively. For a valid comparison, the EEG data in the benchmark database were band-pass filtered between 3 and 100 Hz (eegfilt in EEGLAB) before epoching. Trials in this database were padded with zeros (3 s for S1–S15, and 2 s for S16–S70) to provide a spectral resolution of 0.2 Hz. **Figure 4** illustrates the normalized histogram of the narrow-band (**Figure 4A**) and wide-band SNRs (**Figure 4B**) for the trials in the two databases. For the narrow-band SNR, the BETA database had a significantly lower SNR (3.996 ± 0.018 dB) than the benchmark database (8.157 ± 0.024 dB), with a p-value of < 0.001, $z = -142.212$, Mann-Whitney U-test. Similarly, the wide-band SNR of the BETA database (−13.779 ± 0.013 dB) was significantly lower than the benchmark database (−10.918 ± 0.017 dB), with a p-value of < 0.001, $z = -121.571$, Mann-Whitney U-test. This was due in part to the individual differences in the SNR values of the two studies and in part because the EEG data were recorded outside the electromagnetic shielding room in the BETA database. The comparable results of the two SNR values also demonstrate the validity of the wide-band SNR metric that takes into account additional information on the wide-band noise and harmonics.

In addition, the characteristics of SNR were analyzed with respect to each stimulus frequency. For the BETA database, the wide-band SNRs were calculated for the zero-padded trials, and the SNR associated with each condition was obtained by averaging the values per block and per person. **Figure 5** illustrates the wide-band SNR corresponding to the 40 stimulus frequencies. In general, a declining tendency in SNR can be observed as the stimulus frequency increases. However, at some stimulus frequencies, e.g., 11.6, 10.8, 12, and 9.6 Hz, the SNR bumps up compared to their adjacent frequencies. Specifically, the average SNR value at 15.8 Hz was elevated by 1.49 dB compared to 15.6 Hz, which presumably was due in part to the larger region of visual stimulation.

4.3. Phase and Visual Latency Estimation

In order to further compare the BETA database with the benchmark database in Wang et al. (2017), we estimated the phase and visual latency of the BETA database. Nine consecutive stimulus frequencies in the first row of the keyboard were selected, and the SSVEP from the Oz channel (70 subjects) was extracted for analysis. The comparison procedure was performed according to that in the previous study (Wang et al., 2017) using a linear regression between the estimated phase and stimulus frequency (Russo and Spinelli, 1999). The visual latency for each subject using the slope k of the linear regression is obtained as follows:

$$Latency = -500 \cdot k \tag{6}$$

Figure 6 illustrates the phase as a function of the stimulus frequency, and the bar plot of the estimated latencies estimated by (6). The mean estimated visual latency was 124.96 ± 14.81 ms, which was close to 136.91 ± 18.4 ms of the benchmark database (Wang et al., 2017) and approximated to 130 ms. Therefore, a 130-ms latency was added to the SSVEP epochs for the subsequent classification analysis.

4.4. Accuracy and ITR on Various Algorithms

In this study, 11 frequency recognition methods, including six supervised methods and five training-free methods, were adopted to evaluate the BETA database. For S1–S15, the epoch length of 2 s was used for analysis, and for S16–S70, the epoch length was 3 s. A sliding window from the stimulus onset (latency corrected) with an interval of 0.2 s was applied to the epochs for offline analysis.

4.4.1. Supervised Methods

We choose six supervised methods, including the task-related component analysis (TRCA, Nakanishi et al., 2018), multi-stimulus task-related component analysis (msTRCA, Wong et al., 2020a), Extended CCA (Nakanishi et al., 2014), modified Extended CCA (m-Extended CCA, Chen et al., 2015a), L1-regularized multiway CCA (L1MCCA, Zhang et al., 2013), and individual template-based CCA (ITCCA, Bin et al., 2011) for comparison. The leave-one-out procedure on four blocks was applied to each subject to calculate the accuracy and ITR. **Figure 7** illustrates the average accuracy and the ITR of the supervised methods. The results showed that the msTRCA outperformed other methods at data lengths < 1.4 s, and the m-Extended CCA achieved the highest performance at data lengths from 1.6 to 3 s. The one-way repeated measures ANOVA revealed that there were significant differences between the methods in the ITRs for all time windows. Specifically, for a short time window of 0.6 s, the main effect of methods showed there was a statistically significant difference in ITR, $F_{(1.895,130.728)} = 186.528$, $p < 0.001$, partial $\eta^2 = 0.730$. *Post-hoc* paired t-tests showed that the order was as follows: msTRCA > TRCA > m-Extended CCA > Extended CCA > ITCCA > L1MCCA in ITR, where ">" indicates p was <0.05 in the ITR with Bonferroni correction for pairwise comparison between the two sides. For a medium-length time window of 1.2 s, the main effect of methods showed there was a statistically significant difference in ITR, $F_{(1.797,124.020)} = 197.602$, $p < 0.001$, partial $\eta^2 = 0.741$. *Post-hoc* paired t-tests showed the following: msTRCA / m-Extended CCA / TRCA > Extended CCA > ITCCA > L1MCCA (msTRCA vs m-Extended CCA: $p = 0.678$; m-Extended CCA vs TRCA: $p = 1.000$; Bonferroni corrected). The data length corresponding to the highest ITR varied between different methods; namely, the following results were achieved: msTRCA: 145.26 ± 8.15 bpm at 0.6 s, TRCA: 139.58 ± 8.52 bpm at 0.6 s, m-Extended CCA: 130.58 ± 7.53

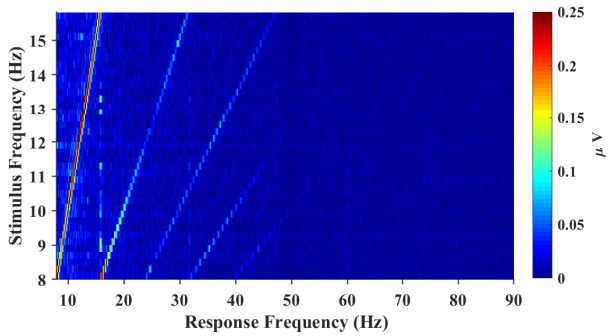


FIGURE 3 | The amplitude spectrum as a function of stimulus frequency (frequency range: 8–15.8 Hz; frequency interval: 0.2 Hz). The spectral response of SSVEP decreases rapidly as the number of harmonics increases and up to 5 harmonics are visible from the figure.

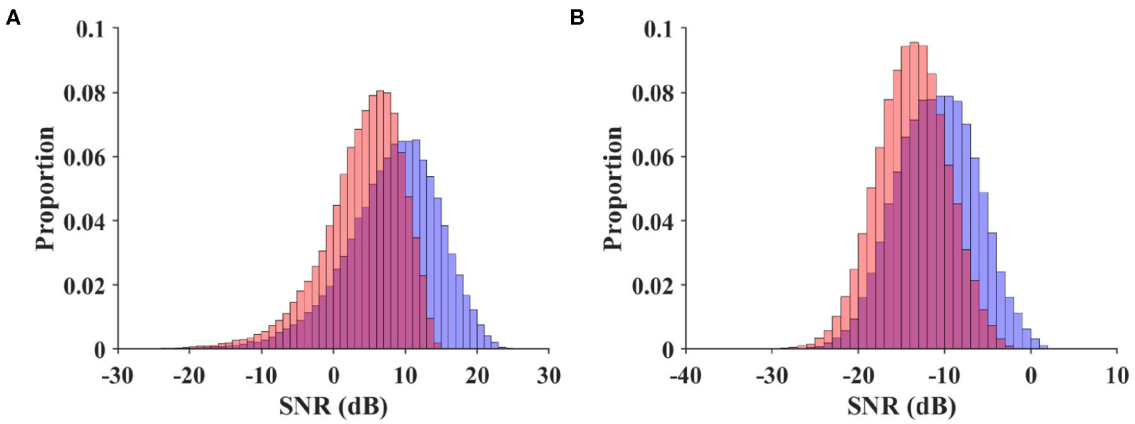


FIGURE 4 | Normalized histogram of narrow-band SNR **(A)** and wide-band SNR **(B)** for trials in the benchmark database and BETA. The red diagram indicates the BETA, and the blue diagram indicates the benchmark database.

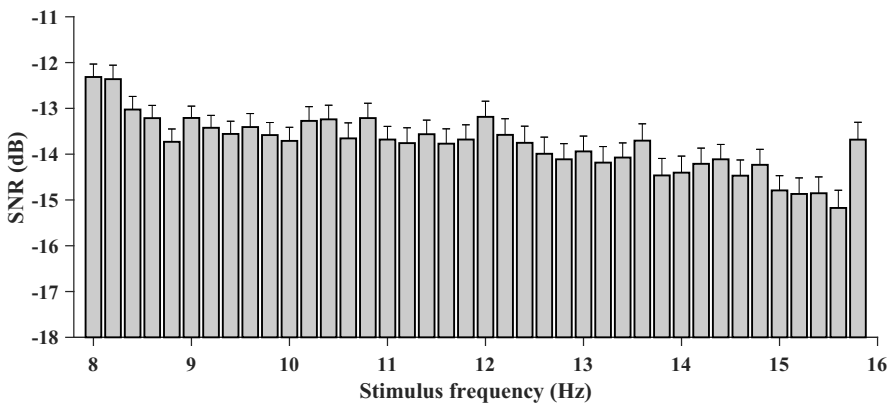


FIGURE 5 | The wide-band SNR corresponding to the 40 stimulus frequencies (from 8 to 15.8 Hz with an interval of 0.2 Hz). A general declining tendency of SNR with the stimulus frequency can be observed. The SNR is higher at 15.8 Hz presumably because the target has a larger shape of the region.

bpm at 0.8 s, Extended CCA: 119.17 ± 6.67 bpm at 1 s, ITCCA: 88.72 ± 6.75 bpm at 1 s, L1MCCA: 73.42 ± 5.31 bpm at 1.4 s).

4.4.2. Training-Free Methods

In this study, five training-free methods, including the minimum energy combination (MEC, Friman et al., 2007), canonical

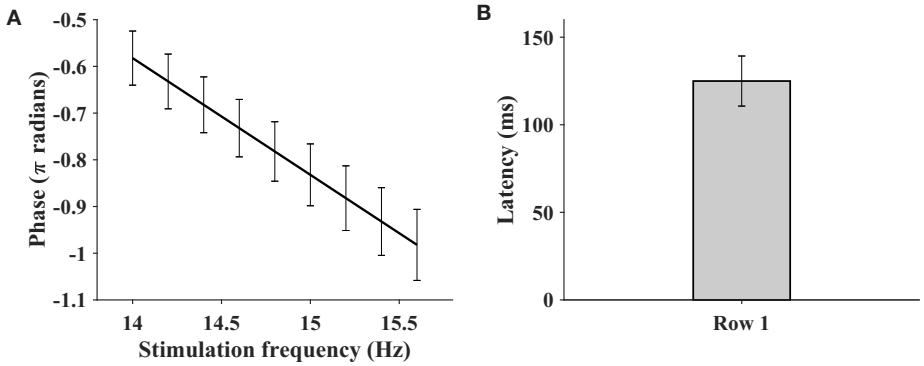


FIGURE 6 | The phase as a function of stimulus frequency **(A)** and the bar plot of estimated latencies **(B)**. The SSVEP of Oz channel at nine consecutive stimulus frequencies (row 1 of the keyboard) is extracted for the purpose of analysis. The error bar indicates the standard deviation.

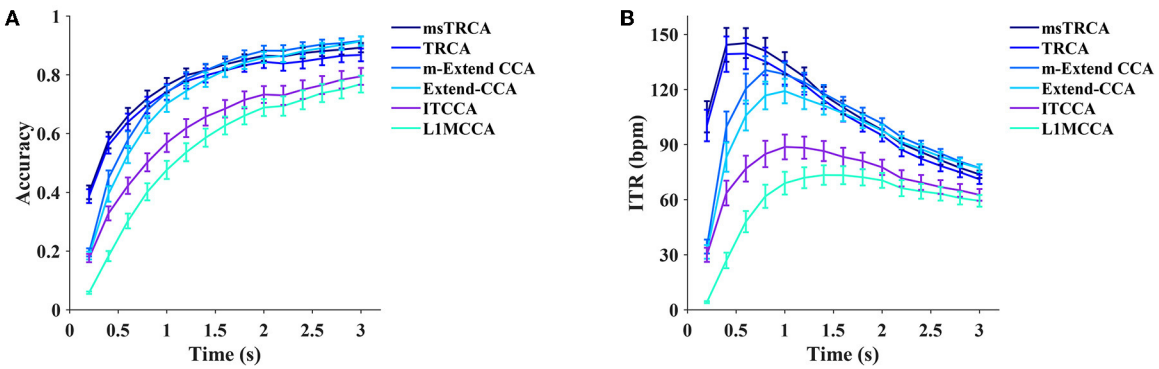


FIGURE 7 | The average classification accuracy **(A)** and ITR **(B)** for six supervised methods (msTRCA, TRCA, m-Extended CCA, Extended CCA, ITCCA, and L1MCCA). Ten data lengths ranging from 0.2 to 3 s with an interval of 0.2 s were used for evaluation. The gaze shift time used the calculation of ITR was 0.55 s.

correlation analysis (CCA, Bin et al., 2009), multivariate synchronization index (MSI, Zhang et al., 2014), filter bank canonical correlation analysis (FBCCA, Chen et al., 2015b), and canonical variates with autoregressive spectral analysis (CVARS, Abu-Alqumsan and Peer, 2016) are compared. As illustrated in **Figure 8**, the FBCCA was superior over the other methods at data lengths <2 s, and the CVARS outperformed the others at data lengths from 2 to 3 s. Significant differences in ITR were found between the methods by the one-way repeated measures ANOVA for all the data lengths. For a medium-length time window of 1.4 s, the main effect of methods showed there was a statistically significant difference in ITR, $F_{(1.876,129.451)} = 79.227$, $p < 0.001$, partial $\eta^2 = 0.534$. *Post-hoc* paired *t*-tests with Bonferroni correction showed the following result: FBCCA > CVARS > CCA / MSI / MEC, $p < 0.05$ for all pairwise comparisons except CCA vs MSI ($p = 1.000$), CCA vs. MEC ($p = 1.000$), MSI vs. MEC ($p = 1.000$). As for the training-free methods, the highest ITR was achieved after 1.2 s, and the result was as follows: FBCCA: 98.79 ± 4.49 bpm at 1.4 s, CVARS: 93.08 ± 4.39 bpm at 1.6 s, CCA: 72.54 ± 4.54 bpm at 1.8 s, MSI: 74.54 ± 4.46 bpm at 1.8 s, MEC: 73.23 ± 4.43 bpm at 1.8 s.

Note that for the TRCA and msTRCA, the ensemble and filter-bank scheme were employed by default. Therefore, to ensure a fair comparison, the number of harmonics N_h was set to 5 in all the methods with sinusoidal templates except the m-Extended CCA according to Chen et al. (2015a) ($N_h = 2$). For all the methods without a filter bank scheme, the trials were band-pass filtered between 6 and 80 Hz except for the CVARS method, which was in line with the previous study (Nakanishi et al., 2015).

4.5. Correlation Between SNR and ITR

In order to explore the relationship between the SNR and ITR metrics, the wide-band and narrow-band SNRs were both investigated. The maximum ITR for each subject (after averaging the ITR values by block) from the training-free FBCCA was chosen for the analysis. **Figure 9** illustrates the scatter plots of the narrow-band and wide-band SNRs vs the ITR. As can be seen in **Figure 9**, the ITR was positively correlated with the SNR for both the narrow-band and wide-band values. For the narrow-band SNR, the statistical analysis reveals that the metric could significantly predict the ITR, $F_{(1,68)} = 45.600, p < 0.001$, and the narrow-band SNR accounted for 40.1% of the variation in the ITR with adjusted $R^2 = 0.393$. The wide-band SNR could also statistically significantly predict the ITR, $F_{(1,68)} = 84.944, p < 0.001$, accounting for 55.5% of the variation in the ITR with adjusted $R^2 = 0.549$. This result indicates that the metric of a

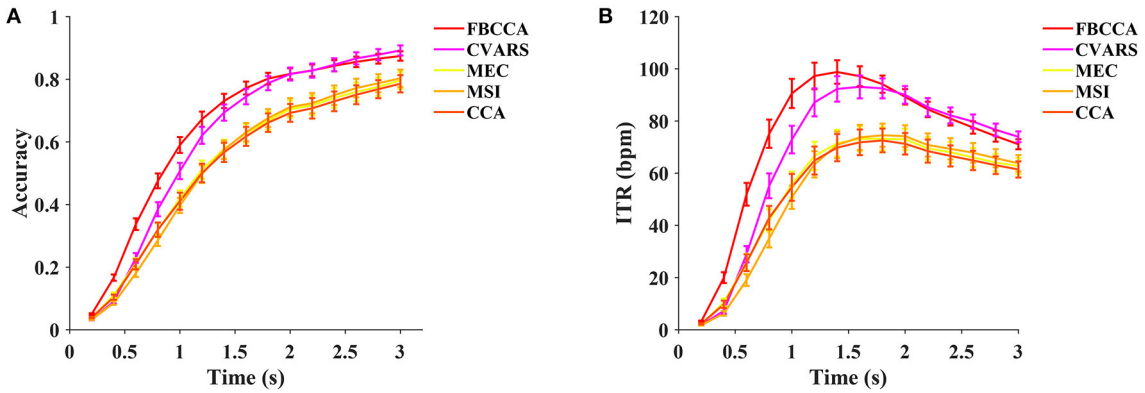


FIGURE 8 | The average classification accuracy **(A)** and ITR **(B)** of five training-free methods (FBCCA, CVARS, MEC, MSI, and CCA). Ten data lengths ranging from 0.2 to 3 s with an interval of 0.2 s were used for evaluation. The gaze shift time used the calculation of ITR was 0.55 s.

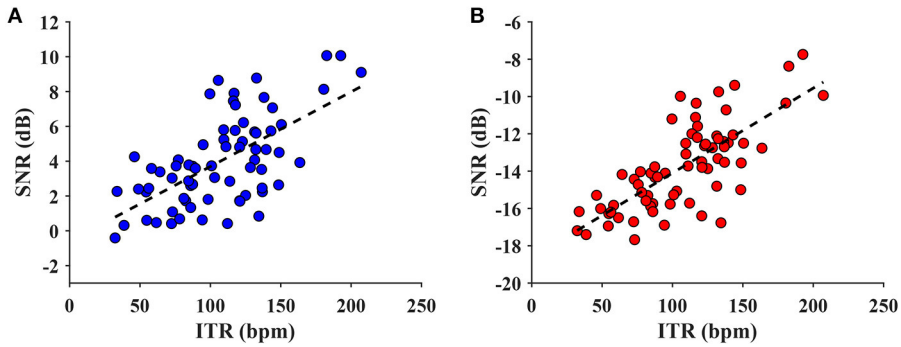


FIGURE 9 | The scatter plot of narrow-band SNR vs. ITR **(A)** and wide-band SNR vs. ITR value **(B)**. The dash line indicates a linear model regressed on the data (**A**: adjusted $R^2 = 0.393$, $p < 0.001$; **B**: adjusted $R^2 = 0.549$, $p < 0.001$). The regression indicates that the wide-band SNR correlated better with the ITR than the narrow-band SNR.

wide-band SNR is more correlated with and can predict better ITR than a narrow-band SNR.

4.6. BCI Quotient

The electroencephalographic signals, including the SSVEP showed individual differences in population. In this study, we propose a BCI quotient to characterize the subject's capacity to use the SSVEP-BCI measured at the population level. Equivalent to the scoring procedure of intelligence quotient (IQ) (Wechsler, 2008), the (SSVEP-) BCI quotient is defined as follows:

$$Quotient_{BCI} = 15 \cdot \frac{SNR - \mu}{\sigma} + 100 \tag{7}$$

where SNR represents the wide-band SNR, and the mean and standard deviation in this study are $\mu = -13.78$ and $\sigma = 2.31$, respectively, as shown in **Figure 10**. The mean and standard deviation can be estimated more accurately for a larger database in the future. The BCI quotient rescales an individual's SNR of the SSVEP to the range of normal distribution $\mathcal{N}(100, 15)$. Since the BCI quotient denotes a relative value derived from SNR, and SNR is correlated with the ITR, the BCI quotient has the potential to measure signal quality and performance for individuals in the SSVEP-BCI. Higher BCI quotient values indicate a higher probability of good BCI performance. For instance, the BCI quotients of S20 and S23 were 74.71 and 139.21, respectively, which reveals a prior to the individual level of the ITR, i.e., 73.09 bpm for S20 and 192.63 bpm for S23. The BCI quotients for each subject were listed in **Table S1** and the result of a regression analysis between the BCI quotient and ITR was provided in the **Supplementary Material**.

5. DISCUSSION

5.1. Data Quality and Its Applicability

Compared to the benchmark database (Wang et al., 2017), the BETA database had lower SNR and the corresponding ITR in the classification (for the benchmark database: FBCCA, 117.96±7.78 bpm at 1.2 s; m-Extended CCA, 190.41 ± 7.90 bpm at 0.8 s; CCA, 90.16 ± 6.81 bpm at 1.6s; 0.55-s rest time for comparison; Chen et al., 2015b; Wang et al., 2017). This can be expected since, in BCI applications, neither there is actually electromagnetic shielding condition nor can be ensured that each subject has

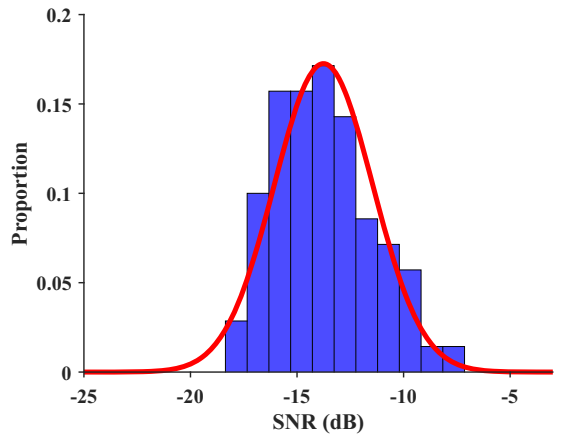


FIGURE 10 | The distribution of wide-band SNR and its fitting to a normal distribution. An individual's SNR of the SSVEP is rescaled to the range of normal distribution by Equation (7) to obtain the BCI quotient.

a high SNR of the SSVEP. The discrepancy in SNR was due in part to the distinct stimulus duration, which was 2 or 3 s for the BETA database and 5 s for the benchmark database. However, even at the same stimulus duration (a 3-s trial after stimulus onset for 55 subjects in the BETA and 35 subjects in the benchmark), the BETA database had significantly lower SNR than the benchmark database (narrow-band SNR: BETA 4.296 ± 0.021 dB, benchmark 5.218 ± 0.020 dB, $p < 0.001$, $z = -34.039$, Mann-Whitney U-test; wide-band SNR: BETA -13.531 ± 0.015 dB, benchmark -12.912 ± 0.015 dB, $p < 0.001$, $z = -28.814$, Mann-Whitney U-test). Therefore, the present BETA database poses challenges to the traditional frequency recognition methods and provides opportunities for the development of robust frequency recognition algorithms intended for real-world applications.

A large number of subjects in the BETA database has the merit of reducing the over-fitting and can provide an unbiased estimation in the evaluation of frequency recognition algorithms. Also, a large volume of the BETA provides an opportunity for the research on transfer learning for the purpose of exploiting common discriminative patterns across subjects. Note that in the BETA database, the number of blocks of each subject is smaller than that in the benchmark database. Since reducing the training and calibration time is critical for the BCI application, the proposed database can serve as a test-bed for the development of supervised frequency recognition methods based on smaller training samples or few-shot learning. It is noteworthy that the application scenario of the BETA database is not limited to the 40-target speller presented in the study. Namely, practitioners can select a subset of the 40 targets (e.g., 4, 8, 12 targets) and design customized paradigms to meet the requirements of a variety of real-world applications. However, since the paradigm of the BETA database falls into the category of dependent BCI where subjects were instructed to redirect their gaze during target selection, the gaze shifting limits its applicability for patient users challenged by oculomotor control. Specifically for these scenarios, gaze independent SSVEP-BCI that is based on covert selective attention (Kelly et al., 2005; Allison et al., 2008; Zhang et al., 2010; Tello et al., 2016) or stimulation via closed eyes (Lim et al., 2013; Hwang et al., 2015) could be deployed, although the information throughput is low with only 2 or 3 targets and modest accuracy. Nevertheless, the BETA database shows its potential to unlock new applications in SSVEP-BCI for alternative and augmentative communication. With the advent of big data, the BETA shows promise for facilitating brain modeling at a population level and help developing novel classification approaches or learning methodology, such as federated learning (Mcmahan et al., 2017) based on big data.

5.2. Supervised and Training-Free Methods

In general, the state-of-the-art supervised frequency recognition methods have the advantage of higher performance regarding the ITR, and the training-free methods excel in ease of use. In this study, two of the supervised methods (the m-Extended CCA, and the Extended CCA) outperformed the five training-free algorithms at all the data lengths. Specifically, at the short-time window (0.2–1 s) the supervised methods (the msTRCA, the TRCA, the m-Extended CCA, and the Extended CCA) outperformed the training-free methods by a large margin (see **Figure S1**). This was because the introduction of the EEG training template and the learned spatial filters facilitated the SSVEP classification. At the time window longer than 2 s (2.2–3 s), the *post hoc* paired t-tests showed that no significant difference is between the m-Extended CCA and the Extended CCA, between the FBCCA and the CVARS, and among the ITCCA, the CCA, the MEC, and the MSI ($p > 0.05$, Bonferroni corrected). Such a result suggests certain common mathematical grounds shared by these algorithms in principle (Wong et al., 2020b). Interestingly, as reported in the previous study (Nakanishi et al., 2018), the TRCA method performance decreased presumably due to the lack of sufficient training block for subjects with low SNR. As evidenced by the previous study (Nakanishi et al., 2018), for the TRCA the number of training data greatly affects classification accuracy ($\approx$ 0.85 with 11 training blocks and $\approx$ 0.65 with two training blocks for a 0.3-s time window). This implies that methods with a sinusoidal reference template (e.g., m-Extended CCA, Extended CCA, and FBCCA, etc.) may be more robust than those without it (Wong et al., 2020b). To sum up, the presented classification analysis demonstrates the utility of different competing methods on the BETA. Besides, the comparison of different methods on a single database complements the previous work of Zerafa et al. (2018), where the performance of various methods was not compared on the same database.

5.3. SNR Comparison

The SNR-based analysis results showed that the wide-band SNR was more correlated with the ITR than the narrow-band SNR. As shown in **Figure 4**, a transition from the narrow-band SNR to the wide-band SNR did not affect the relative relationship between the SNRs of the two databases. Nevertheless, the wide-band SNR metric reduces the skewness of data distribution from -0.708

to −0.081 for benchmark database, and from −1.108 to −0.142 for the BETA database; the narrow-band SNR was followed by the wide-band SNR, which makes the SNR characteristic be more likely to follow the Gaussian distribution. According to Parseval's theorem, the spectral power of a signal is equal to its power in the time domain, so the formulated wide-band SNR has equivalent mathematical underpinning as a metric of temporal SNR counterpart. Apart from its expressive power of wide-band SNR, this metric is also intuitive in the description of signal and noise due to the frequency tagging attribute of the SSVEP.

6. CONCLUSION

In this paper, a **BE**nchmark database **T**owards BCI **A**pplication (BETA) for the 40-target SSVEP-BCI paradigm is presented. The BETA database is featured by its large number of subjects and its paradigm that is well-suited for real-world applications. The quality of the BETA is validated by the typical temporal, spectral and spatial profile of the SSVEP, together with the SNR and the estimated visual latency. The BETA database compares eleven frequency recognition methods, including six supervised methods and five training-free methods. The result of classification analysis validates the data and demonstrates the performance of different methods in one arena as well. As for the metric to characterize the SSVEP, we recommend adopting the wide-band SNR at the single-trial level and use the BCI quotient at the population level. We expect the proposed BETA database can pave the way for the development of methods and paradigms for practical BCI and push the boundary of the BCI toward real-world application.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: http://bci.med.tsinghua.edu.cn/download.html. The datasets has an alternative source for download at https://figshare.com/articles/The_BETA_database/12264401 for the sake of stable access.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Tsinghua University. Written informed consent to participate in this study was provided by the participants, and where necessary, the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

BL conducted the data curation and analysis and wrote the manuscript. XH designed the paradigm and performed the data collection. YW and XC performed the data collection and revised the manuscript. XG supervised the study. All authors contributed to the article and approved the submitted version.

FUNDING

The research presented in this paper was supported by the Doctoral Brain + X Seed Grant Program of Tsinghua University, National Key Research and Development Program of China (No. 2017YFB1002505), Strategic Priority Research Program of Chinese Academy of Science (No. XDB32040200), Key Research and Development Program of Guangdong Province (No. 2018B030339001), and National Natural Science Foundation of China under Grant 61431007.

ACKNOWLEDGMENTS

We would like to thank L Liang, D Xu, J Sun, and X Li for providing support in the data collection, and thank N Shi and J Chen for the manuscript preparation. We would like to thank the reviewers who offered constructive advice on the manuscript. This manuscript has been released as a pre-print at http://arxiv.org/pdf/1911.13045 (Liu et al., 2019).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2020.00627/full#supplementary-material

REFERENCES

Abu-Alqumsan, M., and Peer, A. (2016). Advancing the detection of steady-state visual evoked potentials in brain–computer interfaces. *J. Neural Eng.* 13:036005. doi: 10.1088/1741-2560/13/3/036005

Allison, B. Z., McFarland, D. J., Schalk, G., Zheng, S. D., Jackson, M. M., and Wolpaw, J. R. (2008). Towards an independent brain–computer interface using steady state visual evoked potentials. *Clin. Neurophysiol.* 119, 399–408. doi: 10.1016/j.clinph.2007.09.121

Ang, K. K., and Guan, C (2013). Brain-computer interface in stroke rehabilitation. *J. Comput. Sci. Eng.* 7, 139–146. doi: 10.5626/JCSE.2013.7.2.139

Bakardjian, H., Tanaka, T., and Cichocki, A. (2010). Optimization of SSVEP brain responses with application to eight-command brain–computer interface. *Neurosci. Lett.* 469, 34–38. doi: 10.1016/j.neulet.2009.11.039

Bin, G., Gao, X., Wang, Y., Li, Y., Hong, B., and Gao, S. (2011). A high-speed BCI based on code modulation VEP. *J. Neural Eng.* 8:025015. doi: 10.1088/1741-2560/8/2/025015

Bin, G., Gao, X., Yan, Z., Hong, B., and Gao, S. (2009). An online multi-channel SSVEP-based brain–computer interface using a canonical correlation analysis method. *J. Neural Eng.* 6:046002. doi: 10.1088/1741-2560/6/4/046002

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vis.* 10, 433–436. doi: 10.1163/156856897x00357

Chen, X., Chen, Z., Gao, S., and Gao, X. (2014). A high-ITR SSVEP-based BCI speller. *Brain Comput. Interfaces* 1, 181–191. doi: 10.1080/2326263x.2014.944469

Chen, X., Wang, Y., Gao, S., Jung, T. P., Gao, X. (2015b). Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain–computer interface. *J. Neural Eng.* 12:046008. doi: 10.1088/1741-2560/12/4/046008

Chen, X., Wang, Y., Nakanishi, M., Gao, X., Jung, T. P., and Gao, S. (2015a). High-speed spelling with a noninvasive brain–computer interface. *Proc. Natl. Acad. Sci. U.S.A.* 112, E6058–E6067. doi: 10.1073/pnas.1508080112

Cheng, M., Gao, X., Gao, S., and Xu, D. (2002). Design and implementation of a brain-computer interface with high transfer rates. *IEEE Trans. Biomed. Eng.* 49, 1181–1186. doi: 10.1109/tbme.2002.803536

Choi, G. Y., Han, C. H., Jung, Y. J., and Hwang, H. J. (2019). A multi-day and multi-band dataset for a steady-state visual-evoked potential–based brain-computer interface. *GigaScience* 8:giz133. doi: 10.1093/gigascience/giz133

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Friman, O., Volosyak, I., and Graser, A. (2007). Multiple channel detection of steady-state visual evoked potentials for brain-computer interfaces. *IEEE Trans. Biomed. Eng.* 54, 742–750. doi: 10.1109/tbme.2006.889160

Gao, X., Xu, D., Cheng, M., and Gao, S. (2003). A bci-based environmental controller for the motion-disabled. *IEEE Trans. Neural Syst. Rehabil. Eng.* 11, 137–140. doi: 10.1109/tnsre.2003.814449

Hwang, H. J., Ferreria, V. Y., Ulrich, D., Kilic, T., Chatziliadis, X., Blankertz, B., et al. (2015). A gaze independent brain-computer interface based on visual stimulation through closed eyelids. *Sci. Rep.* **5**:15890. doi: 10.1038/srep15890

Işcan, Z., and Nikulin, V. V. (2018). Steady state visual evoked potential (SSVEP) based brain-computer interface (BCI) performance under different perturbations. *PLoS ONE* 13:e0191673. doi: 10.1371/journal.pone.0191673

Kalunga, E. K., Chevallier, S., Barthélemy, Q., Djouani, K., Monacelli, E., and Hamam, Y. (2016). Online SSVEP-based BCI using riemannian geometry. *Neurocomputing* 191, 55–68. doi: 10.1016/j.neucom.2016.01.007

Kelly, S., Lalor, E., Finucane, C., McDarby, G., and Reilly, R. (2005). Visual spatial attention control in an independent brain-computer interface. *IEEE Trans. Biomed. Eng.* 52, 1588–1596. doi: 10.1109/tbme.2005.851510

Kolodziej, M., Majkowski, A., and Rak, R. J. (2015). "A new method of spatial filters design for brain-computer interface based on steady state visually evoked potentials," in *2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)* (Warsaw). doi: 10.1109/idaacs.2015.7341393

Kwak, N. S., Müller, K. R., and Lee, S. W. (2017). A convolutional neural network for steady state visual evoked potential classification under ambulatory environment. *PLoS ONE* 12:e0172578. doi: 10.1371/journal.pone.0172578

Lebedev, M. A., and Nicolelis, M. A. L. (2017). Brain-machine interfaces: from basic science to neuroprostheses and neurorehabilitation. *Physiol. Rev.* 97, 767–837. doi: 10.1152/physrev.00027.2016

Lee, M. H., Kwon, O. Y., Kim, Y. J., Kim, H. K., Lee, Y. E., Williamson, J., et al. (2019). EEG dataset and OpenBMI toolbox for three BCI paradigms: an investigation into BCI illiteracy. *GigaScience* 8. doi: 10.1093/gigascience/giz002

Lim, J. H., Hwang, H. J., Han, C. H., Jung, K. Y., and Im, C. H. (2013). Classification of binary intentions for individuals with impaired oculomotor function: 'eyes-closed' SSVEP-based brain–computer interface (BCI). *J. Neural Eng.* 10:026021. doi: 10.1088/1741-2560/10/2/026021

Liu, B., Chen, X., Yang, C., Wu, J., and Gao, X. (2017). "Effects of transcranial direct current stimulation on steady-state visual evoked potentials," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Seogwipo).

Liu, B., Huang, X., Wang, Y., Chen, X., and Gao, X. (2019). Beta: A large benchmark database toward SSVEP-BCI application. *aRXiv [preprint]*. arxiv:1911.13045.

Manyakov, N. V., Chumerin, N., Robben, A., Combaz, A., van Vliet, M., Hulle, M. M. V. (2013). Sampled sinusoidal stimulation profile and multichannel fuzzy logic classification for monitor-based phase-coded SSVEP brain–computer interfacing. *J. Neural Eng.* 10:036011. doi: 10.1088/1741-2560/10/3/036011

Mcmahan, H. B., Moore, E., Ramage, D., Hampson, S., and Arcas, B. A. (2017). "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, Vol. 54 (Fort Lauderdale, FL), 1273–1282.

Nakanishi, M., Wang, Y., Chen, X., Wang, Y. T., Gao, X., and Jung, T. P. (2018). Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis. *IEEE Trans. Biomed. Eng.* 65, 104–112. doi: 10.1109/tbme.2017.2694818

Nakanishi, M., Wang, Y., Wang, Y., Mitsukura, Y., and Jung, T. (2014). A high-speed brain speller using steady-state visual evoked potentials. *Int. J. Neural. Syst.* 24:1450019. doi: 10.1142/S0129065714500191

Nakanishi, M., Wang, Y., Wang, Y. T., and Jung, T. P. (2015). A comparison study of canonical correlation analysis based methods for detecting steady-state visual evoked potentials. *PLoS ONE* 10:e0140703. doi: 10.1371/journal.pone.0140703

Nakanishi, M., Wang, Y., Wang, Y. T., and Jung, T. P. (2017). "Does frequency resolution affect the classification performance of steady-state visual evoked potentials?," in *2017 8th International IEEE/EMBS Conference on Neural Engineering (NER)* (Shanghai: IEEE). doi: 10.1109/ner.2017.8008360

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., and Rossion, B. (2015). The steady-state visual evoked potential in vision research: a review. *J. Vis.* 15:4. doi: 10.1167/15.6.4

Pandarinath, C., Nuyujukian, P., Blabe, C. H., Sorice, B. L., Saab, J., Willett, F. R., et al. (2017). High performance communication by people with paralysis using an intracortical brain-computer interface. *Elife* 6:e18554. doi: 10.7554/elife.18554

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015). ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* 115, 211–252. doi: 10.1007/s11263-015-0816-y

Russo, F. D., and Spinelli, D. (1999). Electrophysiological evidence for an early attentional mechanism in visual processing in humans. *Vis. Res.* 39, 2975–2985. doi: 10.1016/s0042-6989(99)00031-0

Tello, R. M., Müller, S. M., Hasan, M. A., Ferreira, A., Krishnan, S., and Bastos, T. F. (2016). An independent-BCI based on SSVEP using figure-ground perception (FGP). *Biomed. Signal Process. Control* 26, 69–79. doi: 10.1016/j.bspc.2015.12.010

Thorpe, S. G., Nunez, P. L., and Srinivasan, R. (2007). Identification of wave-like spatial structure in the SSVEP: comparison of simultaneous EEG and MEG. *Stat. Med.* 26, 3911–3926. doi: 10.1002/sim.2969

van den Broek, S., Reinders, F., Donderwinkel, M., and Peters, M. (1998). Volume conduction effects in EEG and MEG. *Electroencephalogr. Clin. Neurophysiol.* 106, 522–534. doi: 10.1016/s0013-4694(97)00147-8

Wang, Y., Chen, X., Gao, X., and Gao, S. (2017). A benchmark dataset for SSVEP-based brain–computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 1746–1752. doi: 10.1109/tnsre.2016.2627556

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale–Fourth Edition (Wais–IV)*. San Antonio, TX: NCS Pearson.

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain–computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 767–791. doi: 10.1016/S1388-2457(02)00057-3

Wong, C. M., Wan, F., Wang, B., Wang, Z., Nan, W., Lao, K. F., et al. (2020a). Learning across multi-stimulus enhances target recognition methods in SSVEP-based BCIs. *J. Neural Eng.* 17:016026. doi: 10.1088/1741-2552/ab2373

Wong, C. M., Wang, B., Wang, Z., Lao, K. F., Rosa, A., and Wan, F. (2020b). Spatial filtering in SSVEP-based BCIs: unified framework and new improvements. *IEEE Trans. Biomed. Eng.* doi: 10.1109/tbme.2020.2975552. [Epub ahead of print].

Xing, X., Wang, Y., Pei, W., Guo, X., Liu, Z., Wang, F., et al. (2018). A high-speed SSVEP-based BCI using dry EEG electrodes. *Sci. Rep.* 8:14708. doi: 10.1038/s41598-018-32283-8

Yin, E., Zhou, Z., Jiang, J., Yu, Y., and Hu, D. (2015). A dynamically optimized SSVEP brain–computer interface (BCI) speller. *IEEE Trans. Biomed. Eng.* 62: 1447–1456. doi: 10.1109/tbme.2014.2320948

Zerafa, R., Camilleri, T., Falzon, O., and Camilleri, K. P. (2018). To train or not to train? a survey on training of feature extraction methods for SSVEP-based BCIs. *J. Neural Eng.* 15:051001. doi: 10.1088/1741-2552/aaca6e

Zhang, D., Maye, A., Gao, X., Hong, B., Engel, A. K., and Gao, S. (2010). An independent brain–computer interface using covert non-spatial visual selective attention. *J. Neural Eng.* 7:016010. doi: 10.1088/1741-2560/7/1/016010

Zhang, Y., Xu, P., Cheng, K., and Yao, D. (2014). Multivariate synchronization index for frequency recognition of SSVEP-based brain–computer interface. *J. Neurosci. Methods* 221, 32–40. doi: 10.1016/j.jneumeth.2013.07.018

Zhang, Y., Zhou, G., Jin, J., Wang, M., Wang, X., and Cichocki, A. (2013). L1-regularized multiway canonical correlation analysis for SSVEP-based BCI. *IEEE Trans. Neural Syst. Rehabil. Eng.* 21, 887–896. doi: 10.1109/tnsre.2013.2279680

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 12 August 2020
doi: 10.3389/fnins.2020.00849

Analyzing and Decoding Natural Reach-and-Grasp Actions Using Gel, Water and Dry EEG Systems

Andreas Schwarz[1†], Carlos Escolano[2†], Luis Montesano[2,3] and Gernot R. Müller-Putz[1,4]*

[1] Institute of Neural Engineering, Graz University of Technology, Graz, Austria, [2] Bitbrain, Zaragoza, Spain, [3] Departamento de Informática e Ingeniería de Sistemas (DIIS), Instituto de Investigación en Ingeniería de Aragón (I3A), Universidad de Zaragoza, Zaragoza, Spain, [4] BioTechMed Graz, Graz, Austria

Reaching and grasping is an essential part of everybody's life, it allows meaningful interaction with the environment and is key to independent lifestyle. Recent electroencephalogram (EEG)-based studies have already shown that neural correlates of natural reach-and-grasp actions can be identified in the EEG. However, it is still in question whether these results obtained in a laboratory environment can make the transition to mobile applicable EEG systems for home use. In the current study, we investigated whether EEG-based correlates of natural reach-and-grasp actions can be successfully identified and decoded using mobile EEG systems, namely the water-based EEG-Versatile^{TM} system and the dry-electrodes EEG-Hero^{TM} headset. In addition, we also analyzed gel-based recordings obtained in a laboratory environment (g.USBamp/g.Ladybird, gold standard), which followed the same experimental parameters. For each recording system, 15 study participants performed 80 self-initiated reach-and-grasp actions toward a glass (palmar grasp) and a spoon (lateral grasp). Our results confirmed that EEG-based correlates of reach-and-grasp actions can be successfully identified using these mobile systems. In a single-trial multiclass-based decoding approach, which incorporated both movement conditions and rest, we could show that the low frequency time domain (LFTD) correlates were also decodable. Grand average peak accuracy calculated on unseen test data yielded for the water-based electrode system 62.3% (9.2% STD), whereas for the dry-electrodes headset reached 56.4% (8% STD). For the gel-based electrode system 61.3% (8.6% STD) could be achieved. To foster and promote further investigations in the field of EEG-based movement decoding, as well as to allow the interested community to make their own conclusions, we provide all datasets publicly available in the BNCI Horizon 2020 database (http://bnci-horizon-2020.eu/database/data-sets).

Keywords: electroencephalogram, Brain-Computer Interface, reach-and-grasp, movement-related cortical potential, EEG systems, mobile EEG, dry electrodes, BCI data set

OPEN ACCESS

Edited by:
Ana Matran-Fernandez, University of Essex, United Kingdom

Reviewed by:
Imran Khan Niazi, New Zealand College of Chiropractic, New Zealand
Pavel Bobrov, Institute of Higher Nervous Activity and Neurophysiology (RAS), Russia

***Correspondence:**
Gernot R. Müller-Putz
gernot.mueller@tugraz.at

† These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Neural Technology, a section of the journal Frontiers in Neuroscience

Received: *30 May 2020*
Accepted: *21 July 2020*
Published: *12 August 2020*

Citation:
Schwarz A, Escolano C, Montesano L and Müller-Putz GR (2020) Analyzing and Decoding Natural Reach-and-Grasp Actions Using Gel, Water and Dry EEG Systems. Front. Neurosci. 14:849. doi: 10.3389/fnins.2020.00849

INTRODUCTION

The ability to reach-and-grasp is imperative for mastering any actions of daily life and represents the basis of personal independence. It changes for the worse when this ability is taken away, e.g., by a motor vehicle incident, causing a traumatic spinal cord injury (SCI) at cervical level. Needless to say, affected persons, e.g., with a high SCI, seek intervention to regain basic grasping functions (Anderson, 2004; Snoek et al., 2004).

A possible way to regain natural control could be a brain-computer interface (BCI) (Wolpaw et al., 2002; Millán et al., 2010). It enables its users to potentially control any assistive device via voluntary modulation of the users' own brain signals. Brain signals are directly recorded e.g., non-invasively via electroencephalography (EEG) at the scalp of the user and circumvent any damaged parts of the spinal cord. It has been shown that BCIs can be successfully applied for communication (Birbaumer et al., 1999; Kaufmann et al., 2014; Halder et al., 2015; Pinegger et al., 2015; Scherer et al., 2015), however, they can also be used to generate control signals for assistive devices such robotic arms (Meng et al., 2016) or even upper limb motor neuroprostheses (Pfurtscheller et al., 2003; Müller-Putz et al., 2005, 2019; Rohm et al., 2013).

Though control of the designated device could often be successfully established, a major setback was that the control strategies relied on rather abstract mental imaginations and often did not have any direct connection to the intended movement. For instance, Pfurtscheller et al. (2003) relied on the repeated imagery of foot movements and right hand motor imagery to control the study participants' neuroprosthesis attached on the left forearm. We believe that a more natural control strategy is necessary to support an intuitive control for end users (Müller-Putz et al., 2016). Ideally, a future control paradigm consists of one singular non-repetitive task which is similar to the task that has to be performed with such a neuroprosthesis or robotic arm.

Recent investigations have shown that brain patterns of singular upper limb movements can be identified and decoded from EEGs' low frequency time domain (LFTD) signals. These so called movement-related cortical potentials (MRCPs) (Shibasaki et al., 1980) have been shown to hold discriminable information of upper limb movements (Ofner et al., 2017), different grasps (Agashe et al., 2015; Jochumsen et al., 2016), different reach-and-grasp actions (Randazzo et al., 2015; Iturrate et al., 2018; Schwarz et al., 2018, 2019) and can even be decoded online (Ofner et al., 2019; Schwarz et al., 2020).

However, it is still unclear whether the transition from a controlled laboratory environment and its high channel density recording systems to end users' homes utilizing small, mobile EEG systems can be made successfully. The requirements of mobile EEG systems operated at end users' homes are manifold: They need to be (i) easy to handle with the help of a non-expert caregiver and (ii) low in cost and maintenance. From a technical aspect, their (iii) performance needs to be in the same range as their laboratory counterparts, moreover, they (iv) need to operate in a non-laboratory environment. Studies have evaluated usability and performance of emerging mobile systems and compared them to laboratory systems considered "gold standard" (Guger et al., 2012; Pinegger et al., 2016; Di Flumeri et al., 2019). Recently, Jochumsen et al. (2020a,b) evaluated not only the performance of several mobile EEG systems with respect to movement intention detection from the LFTD, but also evaluated their usability with regards to patients, relatives and therapists. Nevertheless, datasets eligible for quantifying different electrode sets are rather scarce.

One of our goals of the Horizon 2020 Project MoreGrasp[1] was to develop a grasp neuroprosthesis for people with SCI which could be operated via a BCI at their homes. As such, we took decisive efforts in designing mobile, state-of-the-art EEG recording systems to provide MoreGrasp end users with BCI technology at their homes (Müller-Putz et al., 2019). Based on these developments, we were able to introduce two market ready recording systems: the water-based electrodes EEG-Versatile^{TM} and the dry electrodes EEG-Hero^{TM}.

The goals of the current study were threefold: Firstly, we wanted to determine whether EEG based correlates of reach-and grasp actions could be extracted. Secondly, we wanted to evaluate whether the LFTD correlates could be successfully decoded and, if so, the potential performance loss due to the transition from a gel-based, gold-standard system to mobile, non gel-based EEG systems. At last, we provide a substantial dataset of 45 study participants recorded with three different EEG systems to the scientific community to foster and promote the research on EEG-based movement decoding.

For this, we assessed the feasibility of the developed recording systems when recording natural reach-and-grasp actions. We performed an experiment in which 45 able bodied participants performed self-initiated reach-and-grasp actions on objects of daily life. Fifteen participants were measured using the mobile and water-based electrodes EEG-Versatile^{TM} system and 15 using the dry-electrodes EEG-Hero^{TM} headset in an office environment. In addition, we provide the recordings of additional 15 able bodied study participants who used a gel based (gold standard) system, who performed the same tasks in a laboratory environment.

MATERIALS AND METHODS

Participants and Recordings

In total, 45 participants took part in the experiment. They were able-bodied and right handed. All gave written informed consent and received monetary compensation for their participation.

Gel-Based Electrodes Recordings

This study was approved by the Medical University of Graz (EK: 30-439 ex 17/18). Recordings using the gel-based recording system (g.tec USBamp/g.tec Ladybird system, g.tec medical engineering GmbH, Austria) were performed at the Institute of Neural Engineering at Graz University of Technology (see **Figure 1**, left). We measured EEG of 15 able bodied, right handed, study participants (10 male, 5 female, aged between 15 and 30, median 26 years) with 58 active electrodes positioned over frontal, central, and parietal areas according to the 5% grid system provided by Oostenveld and Praamstra (2001). Furthermore we recorded the electrooculogram (EOG) using six additional electrodes positioned infra and superior orbital to the left and right eye and on the outer canthi. For reference we used the right earlobe and for ground the channel AFz. All signals were recorded using a sampling frequency of 256 Hz and prefiltered

[1] www.moregrasp.eu

FIGURE 1 | Experimental setup for the three recording systems. **Left** image shows the recordings using the gel-based system (g.tec USBamp/g.tec Ladybird system, g.tec medical engineering GmbH, Austria). **Center** image shows the water-based EEG-Versatile^{TM} (Bitbrain, Spain). **Right** image shows the dry-electrodes EEG-Hero^{TM} system (Bitbrain, Spain). Upper right corner shows the electrodes layout. EOG was recorded in the gel- and water-based electrodes systems (six additional electrodes positioned infra and superior orbital to the left and right eye and on the outer canthi).

using an 8th order Chebyshev filter from 0.01 to 100 Hz. We used a notch filter at 50 Hz to suppress the power noise. All data was synchronized using the TOBI Signal server (Breitwieser et al., 2010). The gel based recordings provided for this study were part of another extended study incorporating also bimanual reach-and-grasp actions, which are not part of the current study. Further details can be found in Schwarz et al. (2019). Force-sensing resistor (FSR) sensors were used to record the movement onset and the grasping time point to each object. Sensor output was digitized using a battery operated Arduino microcontroller.

Water-Based Electrodes Recordings (EEG-Versatile^{TM})

The recordings with the EEG-Versatile^{TM} system (Bitbrain, Spain) were conducted in the office environment of Bitbrain (Zaragoza, Spain), guided by personnel of the Institute of Neural Engineering, Graz University of Technology. We measured the EEG of 15 able bodied, right handed study participants (aged between 15 and 30, median 24 years; 8 females) using 32 water-based electrodes positioned over frontal, central and parietal positions (see **Figure 1**, center). Additionally we used six electrodes positioned infra and superior orbital and the outer canthi to measure EOG. For reference we used the left earlobe and for ground the channel AFz. These signals were recorded using a sample frequency of 256 Hz and prefiltered using a 3rd order anti-aliasing Butterworth filter with pass-band frequency from DC to 100 Hz. Photodiode sensors were used to record the movement onset and the grasping time point to each object. The three photodiodes were digitized using a Biosensing^{TM} amplifier (Bitbrain, Spain) at a sampling rate of 256 Hz, which was placed on the table. Time synchronization between the EEG-EOG signals and photodiodes was made via a TTL output of the Biosensing^{TM} amplifier. All data was streamed via Bluetooth to the computational unit using Bitbrain proprietary software, and backed-up to an internal SD card to avoid data loss due to the wireless connection.

Dry Electrodes Recordings (EEG-Hero^{TM})

The recordings with the EEG-Hero^{TM} headset (Bitbrain, Spain) were also performed in the office environment of Bitbrain (Zaragoza, Spain), guided by personnel of the Institute of Neural Engineering, Graz University of Technology (see **Figure 1**, right). We measured 15 able bodied, right handed study participants (aged between 15 and 30, median 27 years; 7 females) using 11 dry electrodes located over sensorimotor areas according to the international 10/20 system (FC3, FCz, FC3, C3, C1, Cz, C2, C4, CP3, CPz, CP4). For reference and ground we used the left earlobe. These signals were recorded using a sample frequency of 256 Hz and prefiltered using a 3rd order anti-aliasing Butterworth filter with pass-band frequency from DC to 100 Hz. Photodiode sensors were used to record the movement onset and the grasping time point to each object. The three photodiodes were digitized using a Biosensing^{TM} amplifier (Bitbrain, Spain) at a sampling rate of 256 Hz, which was placed on the table. All data was streamed via Bluetooth to the computational unit using Bitbrain proprietary software, and backed-up to an internal SD card to avoid data loss due to the wireless connection.

Experimental Setup and Paradigm

All recordings were performed using the same experimental setup and followed closely the approach presented in Schwarz et al. (2019). However for gel based recordings, the experiment took place in a laboratory environment, where participants were seated in a noise and electromagnetically shielded room. For water-based and dry-electrode based recordings, the experiment took place in a non-shielded office room. Participants were seated on a chair in front of a table and instructed to rest their right hand on a sensorized base position which was positioned in front of them. On the table, we placed an empty jar and a jar with a spoon stuck in it. Both objects were in a comfortable reaching distance equidistant to the study participants' right hand. Participants were instructed to perform reach-and-grasp actions using their right hand towards the objects placed on the table. In case of the empty jar they grasped the objects using a palmar grasp. In case of the spoon, they were instructed to grasp the spoon with a lateral grasp. Though participants performed the tasks in a self-initiated manner, we instructed them to focus their gaze on the designated object for 2 s before initiating the reach-and-grasp action. Once they completed the grasp, they held the object for at least 1–2 s (see **Figure 2**). When they returned their hand to the starting

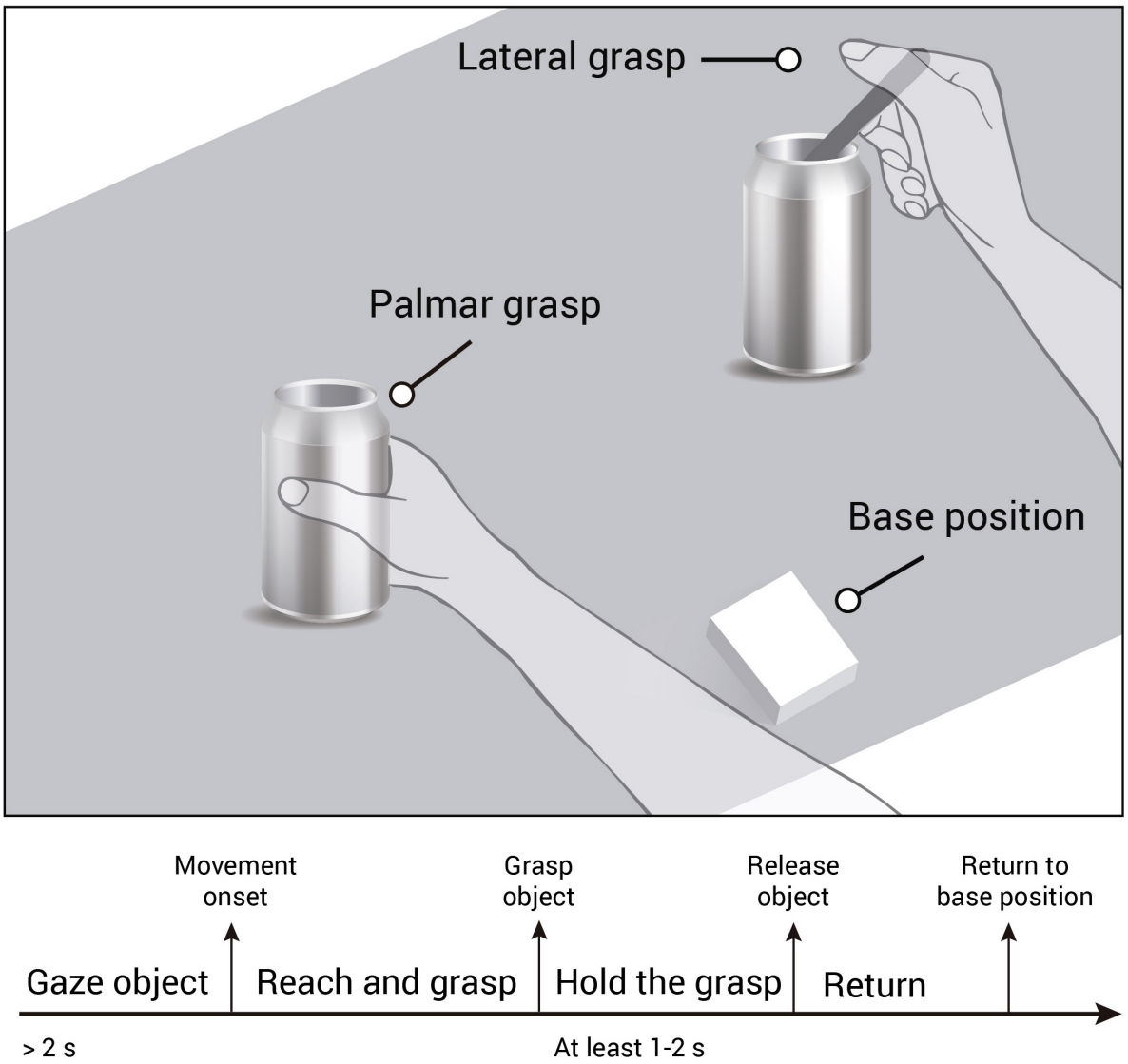


FIGURE 2 | Experimental paradigm and trial timeline. Participants were instructed to gaze at the object for at least 2 s. Then they performed the reach-and-grasp action toward a jar (palmar grasp) or a jar with a spoon stuck in it (lateral grasp), and hold the grasp for at least 1–2 s. Then they returned the hand to the sensorized base position and prepared for the next trial.

position, a small insert on a screen showed them the number of grasps they had already performed on the designated object. In case of the gel-based recordings, the screen was integrated in the table, for the other recording sessions, the screen was positioned in front of them. Lastly, participants paused at least for 4 s before starting a new trial (inter trial interval).

In this way we recorded 80 trials per condition (TPC) distributed over 4 runs á 20 trials. After each run, we switched the position of the objects presented on the table, so that each object was on each position equally.

We also recorded 3 min of rest at the start, after the second movement run (at half time) and at the end of the experiment, where participants were tasked to focus their gaze on a fixation point in the middle of the table. In addition, we recorded horizontal and vertical eye movements as well as blinks following the paradigms used in (Kobler et al., 2018; Schwarz et al., 2019).

Data Preprocessing and Artifact Handling

We filtered all available data using a zero-phase 4th order Butterworth bandpass filter with a cut-off frequency of 0.3 and 60 Hz. For gel-based and water-based recordings, we used all available EEG and EOG channels and applied the extended infomax ICA algorithm on the data. We removed components associated with eye movements and blinks by visual inspection (Lee et al., 1999; Delorme and Makeig, 2004). Note that we refrained to apply an ICA algorithm on the dry-electrode based

recording due to the unfavorable number of channels available ($n = 11$).

We defined a window of interest (WOI) for each movement trial of [−2 3] s with respect to the movement onset at second 0. In addition, we extracted 81 trials from the rest recordings. The rest trials had a duration of 5 s (i.e., similar to the duration of the movement trials).

Subsequently we rejected potentially artifact contaminated data by statistical parameters (Faller et al., 2012; Schwarz et al., 2015, 2018). For each participant's data set, regardless of the recording system, we filtered the data between 0.3 and 35 Hz. We rejected trials by (1) amplitude threshold (amplitude exceeds 125 μV), (2) abnormal joint probability, and (3) abnormal kurtosis by threshold of four times the standard deviation. Trials marked for rejection were excluded from subsequent analysis. As a result, the following trials were rejected for each system: Gel-based sensors (16.0%), Water-based sensors (13.2%), and Dry sensors (8.5%).

Power Spectral Density and Time Frequency Analysis

For each study participant, we applied a common average reference (CAR) filter on the preprocessed EEG data.

For calculating the power spectral density (PSD) estimates, we epoched all trials from (0 1.5) s with respect to the movement onset. Using Welch's method of an overlapping segment averaging estimator, we calculated the PSD using a 1 s window and 25% overlap. We calculated the PSD average per condition and a confidence interval using non-parametric t-percentile bootstrap statistics (alpha = 0.05). To obtain the grand-average PSD, we calculated the mean over the participant-specific average and its respective confidence intervals.

For the time frequency analysis we calculated event-related (de)synchronization (ERD/S) maps in the range from 2 to 40 Hz (1 Hz resolution) as shown by Graimann et al. (2002). The analysis was performed for each movement condition separately using a specific reference interval of (−2 −1) s with respect to the movement onset. To obtain grand-average ERD/S maps (Pfurtscheller and Lopes da Silva, 1999), we calculated for each frequency bin the mean over the participants ERD/S time points and calculated confidence intervals using non-parametric t-percentile bootstrap statistics (alpha = 0.05). The resulting ERD/S maps show only the significant time-frequency points per recording system.

Movement-Related Cortical Potentials

We resampled all preprocessed EEG signals to 16 Hz to save computational load and applied a CAR filter. Thereafter, we applied a 4th order, zero-phase Butterworth lowpass filter with a cut-off frequency of 3 Hz. To allow meaningful comparison across study participants we introduced a normalization step: For each participant, we calculated the global field power (GFP) as the standard deviation across all channels and normalized all scalp potentials by the average GFP of the rest condition (Skrandies, 1990). We epoched all movement trials and the rest recordings according to the WOI (−2 3) s and calculated condition specific averages. In addition, we determined a 95% confidence interval for each condition using non-parametric t-percentile bootstrap statistics. We accumulated a grand average per EEG system by calculating the mean over the participant-specific averages.

Multiclass Single-Trial Classification

The classification approach follows closely the approach presented in Schwarz et al. (2019) and is adapted to the current data set. We resampled all preprocessed EEG signals to 16 Hz to save computational load and applied a CAR filter. Thereafter, we applied a 4th order, zero-phase Butterworth lowpass filter with a cut-off frequency of 3 Hz.

For each study participant, regardless of the recording system, we divided all preprocessed trials of the movement conditions as well as the rest condition in a calibration set, which consisted of the first 66% of all recorded TPC and an unseen test set consisting of the remaining 34% of all recorded TPC.

Using the calibration data set, we assessed the best time point in terms of classification accuracy for training a classification model within the WOI. For that we used a 10 times five fold cross validation approach to divide the calibration set into training and evaluation sets. For each time point within the WOI, we trained an individual shrinkage based linear discriminant classification model (sLDA) (Blankertz et al., 2011). As features, we took nine amplitude values of all available EEG channels (Gel: 58, Water: 32, Dry: 11), from the preceding second of the actual time point in causal steps of 0.125 s (−1:0.125:0) s. This yielded in total 522 features (9 × 58 channels) for the gel-based setup, 288 features (9 × 32 channels) for the water-based setup, and 99 (9 × 11 channels) for the dry electrodes setup. This classification approach was applied on each time point within the WOI yielding in 80 classification models (16 time points × 5 s WOI). To assess the best training time point, we averaged the performance results of all calculated folds and chose the time point with the best average performance. The adjusted chance level was at 45.8% [adjusted Wald interval, alpha = 0.05, (Breitwieser et al., 2012; Müller-Putz et al., 2008)] and corrected for multiple comparisons ($n = 80$ time points) using Bonferroni correction.

Thereafter, we applied the best performing classification model on previously unseen test data, using the same preprocessing pipeline as before. In this case, the adjusted chance level for the test set lies at 42.8% (adjusted Wald interval, alpha = 0.05). We show the mean classification performance of all trials over the whole WOI.

In addition, we repeated this classification approach for gel- and water-based recording systems with a subset of 11 channels covering the sensorimotor electrode positions of the dry electrodes system (EEG-Hero^{TM}).

RESULTS

Gel Based Recordings

Figure 3 (left) shows the grand average time-frequency maps of the gel-based recordings for channels C3, Cz, and C4 with respect to a reference interval (−2 −1) s prior to the movement onset. Significant ERD can be seen on the three channels,

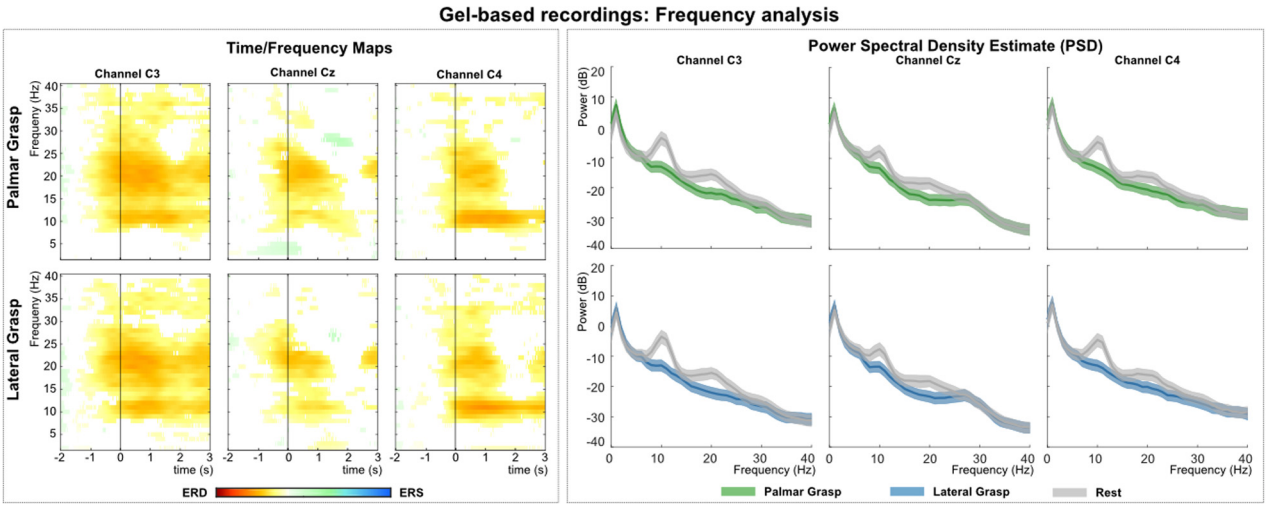


FIGURE 3 | Frequency analysis of the gel-based recordings. **(Left)**: Grand average of the time-frequency maps (Graimann et al., 2002) for movement conditions for positions C3, Cz, and C4 with respect to the reference period (–2 –1) s. The black vertical line represents the movement onset. Hot colors show significant ERD (cold colors represent significant relative increase in power [event-related synchronization (ERS)]. Significant differences with respect to the reference period were calculated using non-parametric t-percentile bootstrap statistics (alpha = 0.05). **(Right)**: Grand average of the PSD calculation of the reaching phase [0 1.5] s. Colored lines represent the PSDs of the movement conditions, gray lines show the PSD of the rest condition. The shaded areas show 95% confidence intervals which were calculated using non-parametric t-percentile bootstrap statistics (alpha = 0.05).

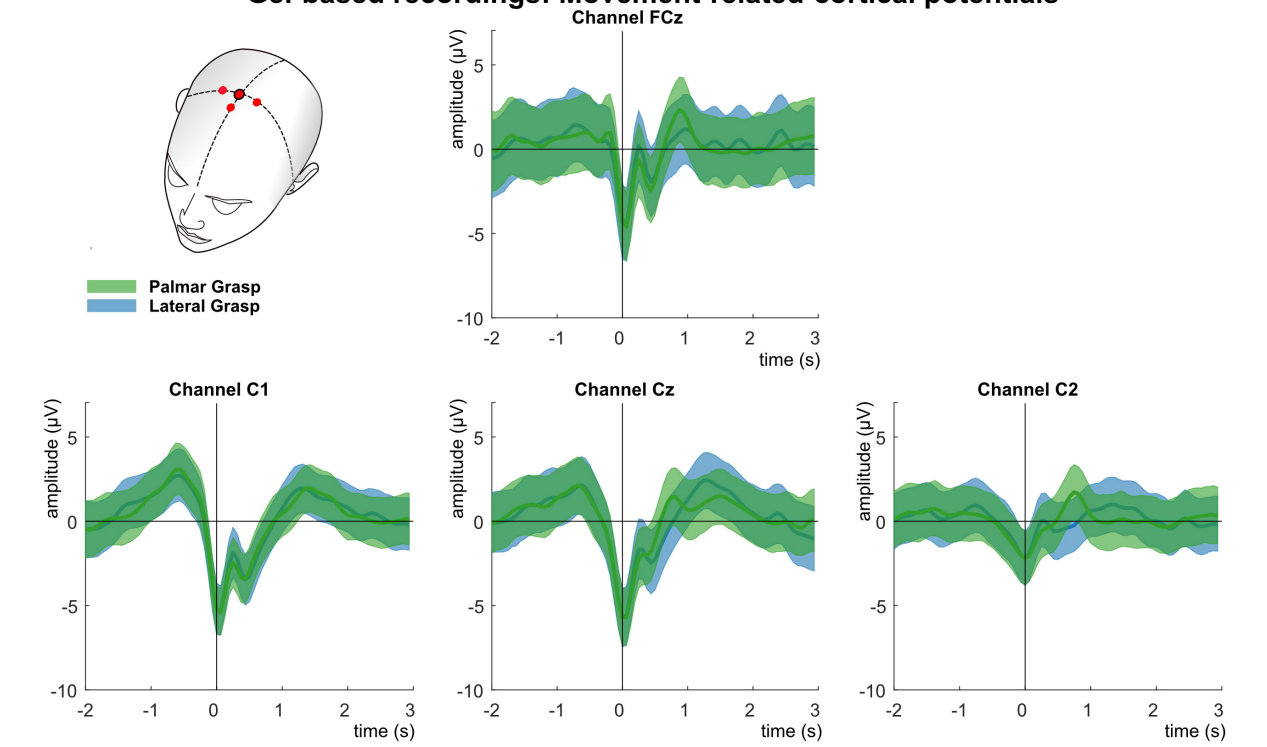


FIGURE 4 | Movement related cortical potentials (MRCPs) of the gel-based recordings. Grand averages (bold lines) and 95% confidence intervals (shaded areas) for palmar (green) and lateral (blue) grasp conditions. Channels shown are FCz, C1, Cz, and C2. The black perpendicular line represents the movement onset.

most prominent in alpha (8–12 Hz) and beta frequency range (∼20 Hz). Alpha (mu) activity shows the most prominent ERD on the ipsilateral side to the executing hand, whereas beta activity shows the most prominent ERD on the contralateral side. **Figure 3** (right) shows the PSD estimates of the reaching phase (0 1.5) s for channels C3, Cz, and C4 for both the movement conditions and the rest condition. When comparing the movement conditions to the rest condition, significant power decreases can be observed mainly in the alpha and beta range. This power decrease is stronger on the contralateral side to the executing hand, especially for the beta frequency range.

Analysis of the MRCPs in the LFTD (see **Figure 4**) shows the grand average for the palmar and lateral grasp conditions. A negative deflection (Bereitschaftspotential, Shibasaki et al., 1980) can be observed (time = 0 s), which starts up to one second before the movement onset. This deflection is pronounced strongest first over the central motor cortex at channel position Cz, and contralateral to the executing right hand second. About 300 ms after the movement onset, a positive deflection (reafferent potential) can be observed. Thereafter, around 1 s after the movement onset, a second positive peak occurs before the potential returns to baseline. On group level, no significant differences between movement conditions could be observed.

Figure 5 summarizes the results of the single trial multiclass decoding of both movement conditions and the rest condition (gel-based recordings). **Figure 5** (top) shows the grand average

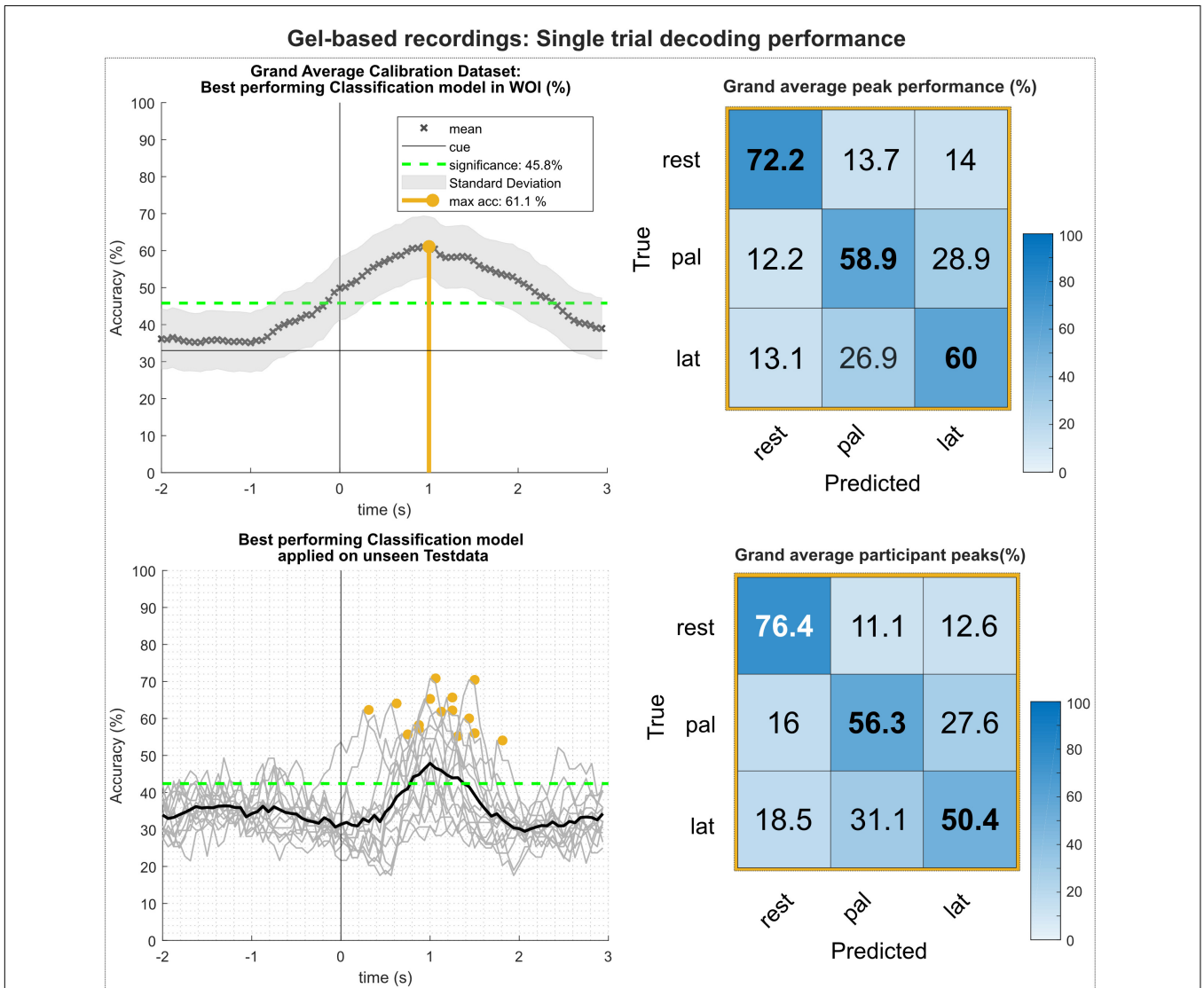


FIGURE 5 | Single trial decoding performance of the gel-based recordings. **(Top left)** Grand average for the calibration data set. Black crosses show the mean performances for the designated time point. On average, best classification performance could be reached 1 s after the movement onset (perpendicular golden line). **(Top right)** Row wise normalized confusion matrix for grand average peak performance. **(Bottom left)** Participant-specific classification results (gray lines) and grand average (black bold line) of the best performing classification model applied on the unseen test data set. Golden dots show participant-specific peak performances. **(Bottom right)** Row wise normalized confusion matrix for the individual peak performances.

obtained on the calibration data set and the corresponding confusion matrix of the grand average peak performance. On average, highest calibration performance could be reached about 1 s after the movement onset with an average peak accuracy of 61.1%. This is lower than the average of the participant-specific (63.9%) peak performance, since the time of peak performance varies between participants. Analysis of the confusion matrix shows that the true positive rate (TPR) for the rest condition is highest with 72.2%, exceeding the TPRs of the movement conditions by more than 12%. In contrast, false positive rates (FPR) for movement versus movement conditions are between 26 and 29%, exceeding FPRs of movement versus rest conditions more than twice. **Figure 5** (bottom) shows the results of the participant-specific best performing classification models applied on the previously unseen test data set and its corresponding confusion matrix of the grand average of the peak performances. Participant-specific peak performances reached on average 61.3% around 1.1 s after the movement onset. The corresponding confusion matrix shows again higher TPRs for classification of the rest versus movement conditions and are within the same range as the results of the calibration data set. However, TPRs for movement versus movement conditions decreased, especially for the lateral grasp condition. **Table 1** depicts the participant-specific classification results.

Water Based Recordings

Figure 6 (left) shows the grand average time-frequency maps of the water-based recordings for channels C3, Cz, and C4 with respect to a reference interval (−2 −1) s prior to the movement onset. Significant ERD can be found for the three channels, especially in alpha (8–12 Hz) and beta (∼20 Hz) band frequencies. The differences are pronounced weakest at central electrode position Cz. Alpha (mu) activity shows the most prominent ERD on the ipsilateral side (C4), with the strongest beta at bilateral positions (C3 and C4). **Figure 6** (right) shows the PSD estimates of the reaching phase (0 1.5) s for channels C3, Cz, and C4 for the movement conditions and the rest condition. Significant differences between movement conditions and rest condition can be observed on channels in the alpha band and on the contralateral side (C3 location) in the beta band.

Figure 7 shows the analysis of the MRCPs for channels FCz, C1, Cz, and C2. Around 1 s before the movement onset, the negative deflection of the BP starts and peaks around movement onset (time = 0 s). The BP is strongest pronounced over the central electrode position first, and contralateral to the executing right hand second. It is clearly recognizable the reafferent potential around 300 ms after the movement onset followed by a second positive deflection around 1–1.5 s after the movement onset before the potentials return back to baseline. The morphologies of both movement conditions are similar and bear no significant difference on group level.

Figure 8 summarizes the results of the multiclass single trial decoding. **Figure 8** (top) shows the grand average obtained on the calibration data set and the corresponding confusion matrix of the grand average peak performance. Grand average peak performance reached 63.6% around 0.9 s after the movement onset. Participant-specific peak accuracies were slightly higher with 65.4%. The confusion matrix shows high TPRs for rest versus movement conditions, exceeding TPRs for movement versus movement conditions by 20%. **Figure 8** (bottom) depicts the participant-specific classification results when applying the best performing classification model trained on the calibration data on the unseen test set. Participant-specific peak performance reaches on average 62.3% around 0.9 s after the movement onset. The corresponding confusion matrix shows an even more favorable TPR for rest versus movement conditions with 81.4%. However, TPRs for movement versus movement conditions decreased, especially for the palmar grasp condition. **Table 2** depicts the participant-specific classification results for calibration and test set in detail.

TABLE 1 | Participant-specific classification results of the gel-based recordings.

	Calibration set			Test set		
#	Peak (%)	STD (%)	Time (s)	Peak (%)	STD (%)	Time (s)
G01	71.2	7.6	0.2	62.3	8.2	0.3
G02	57.0	8.6	1.1	65.7	8.9	1.2
G03	57.0	7.6	1.7	54.1	7.5	1.8
G04	73.6	8.0	0.9	57.1	8.6	0.8
G05	59.9	6.8	1.3	62.2	10.4	1.2
G06	56.4	7.6	0.7	55.7	7.1	0.7
G07	53.6	8.6	0.8	56.0	7.5	1.4
G08	71.6	6.7	1.4	70.4	10.0	1.4
G09	72.0	7.9	1.1	61.8	8.9	1.1
G10	72.0	6.5	0.9	65.3	8.6	0.9
G11	60.3	9.4	0.8	58.2	6.2	0.8
G12	64.9	7.6	1.0	70.8	10.3	1.0
G13	63.8	10.1	1.4	55.2	8.7	1.3
G14	58.1	9.7	0.4	64.1	9.7	0.6
G15	66.5	7.8	1.4	60.0	7.9	1.4
Average	63.9	8.0	1.0	61.3	8.6	1.1

Dry Electrode Recordings

Figure 9 (left) shows the grand average time-frequency maps of the dry electrodes recordings for channels C3, Cz, and C4 with respect to a reference interval (−2 −1) s prior to the movement onset. Significant ERD can be found on the three channels, especially in alpha (8–12 Hz) and beta band (∼20 Hz). Differences are weakest on central channel location Cz. Alpha (mu) activity shows the stronger ERD on the ipsilateral side (C4), and beta presents a stronger ERD on bilateral locations (C3 and C4). **Figure 9** (right) depicts the PSD estimates of the reaching phase [0 1.5] for channels FCz, C1, Cz, and C2. When looking at both movement conditions against the rest condition, a significant power decrease for the movement conditions in alpha and beta can be observed. This power decrease is more pronounced bilaterally than on the central electrode position Cz.

Figure 10 shows the analysis of the MRCPs for channels FCz, C1, Cz, and C2. A negative deflection can be observed at movement onset (time = 0 s), which starts about 0.5 s before the movement onset. It is strongest on the central position first and on the contralateral side to the executing hand second. Around

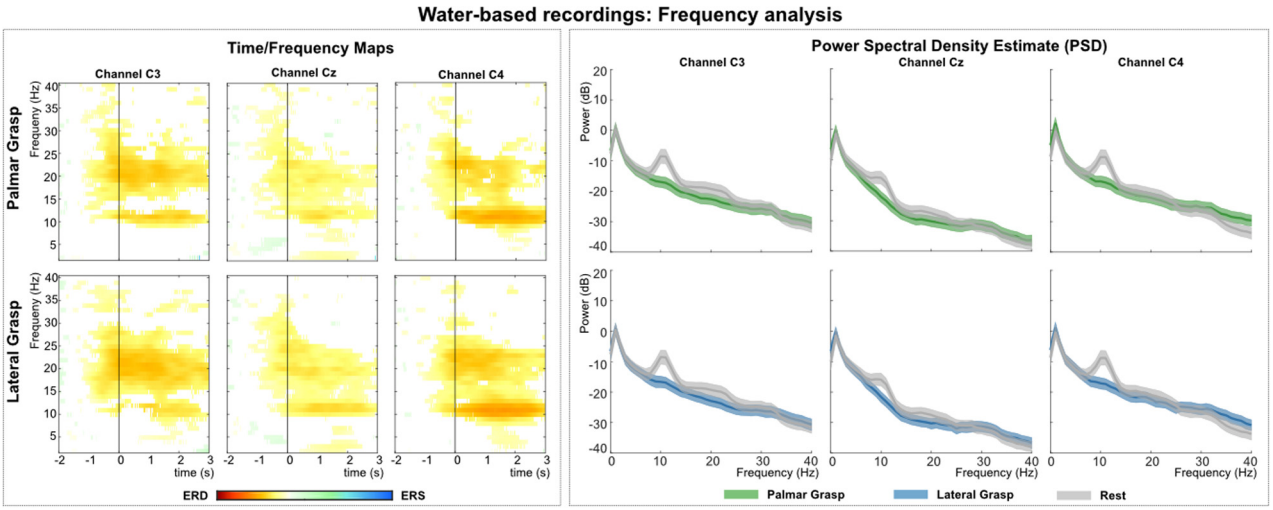


FIGURE 6 | Frequency analysis of the water-based recordings. **(Left)**: Grand average of the time-frequency maps (Graimann et al., 2002) for movement conditions for positions C3, Cz, and C4 with respect to the reference period (–2 –1) s. The black vertical line represents the movement onset. Hot colors show significant ERD [cold colors represent significant relative increase in power (event-related synchronization (ERS)]. Significant differences with respect to the reference period were calculated using non-parametric t-percentile bootstrap statistics (alpha = 0.05). **(Right)**: Grand average of the PSD calculation of the reaching phase [0 1.5] s. Colored lines represent the PSDs of the movement conditions, gray lines show the PSD of the rest condition. The shaded areas show 95% confidence intervals which were calculated using non-parametric t-percentile bootstrap statistics (alpha = 0.05).

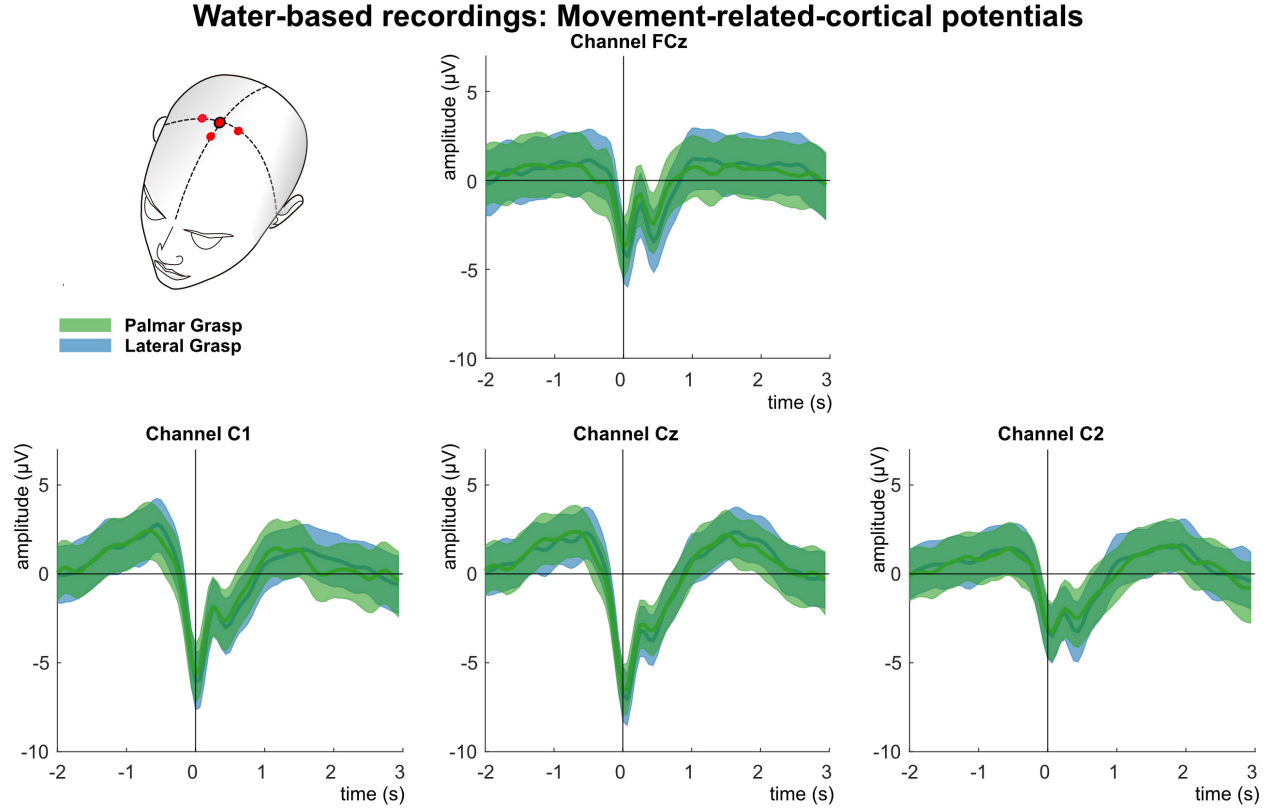


FIGURE 7 | Movement related cortical potentials (MRCPs) of the water-based recordings. Grand averages (bold lines) and 95% confidence intervals (shaded areas) for palmar (green) and lateral (blue) grasp conditions. Channels shown are FCz, C1, Cz, and C2. The black perpendicular line represents the movement onset.

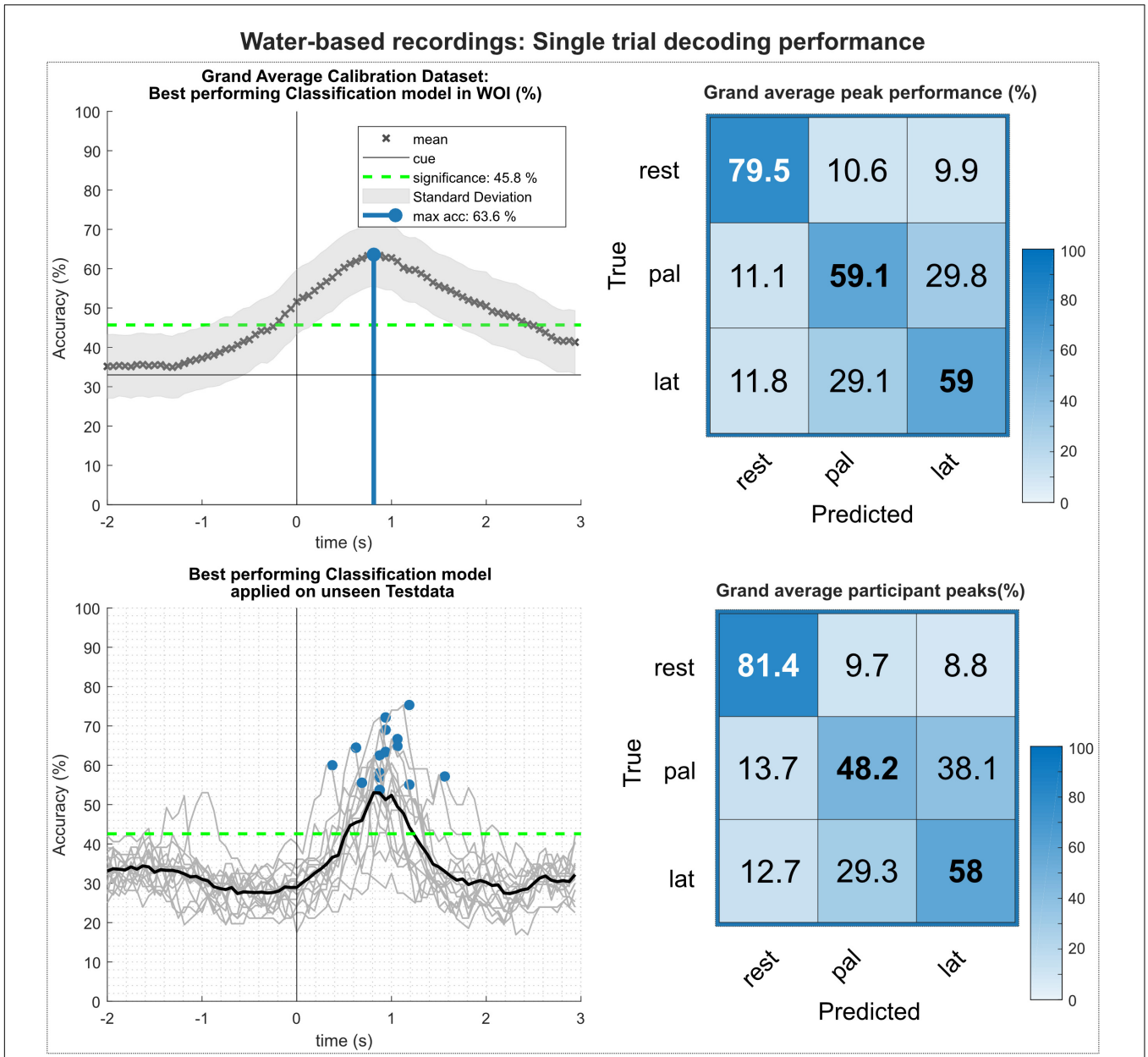


FIGURE 8 | Single trial decoding performance of the water-based recordings. **(Top left)** Grand average for the calibration data set. Black crosses show the mean performances for the designated time point. On average, best classification performance could be reached 0.9 s after the movement onset (perpendicular blue line). **(Top right)** Row wise normalized confusion matrix for grand average peak performance. **(Bottom left)** Participant-specific classification results (gray lines) and grand average (black bold line) of the best performing classification model applied on the unseen test data set. Blue dots show participant-specific peak performances. **(Bottom right)** Row wise normalized confusion matrix for the individual peak performances.

300 ms after the movement onset, at least for electrode positions Cz and C2, a reafferent potential can be seen before the potential returns back to base level about 1–1.5 s after the movement onset. The morphologies of both movement conditions show no significant difference on group level.

Figure 11 shows the results of the multiclass single trial decoding. **Figure 11** (top) depicts the grand average performance of the calibration data set and the confusion matrix of the participant specific grand average peak performance. On average, 56.6% around 1 s after the movement onset could be reached. The participant-specific peak accuracy yielded at 58.3% and is higher due to the variation in timing of reaching peak performance of the participants. The corresponding confusion matrix shows increased TPRs for rest versus movement conditions of 67.5%, whereas TPRs for movement versus movement conditions yielded about 54%. When applying the best performing classification model on the unseen data set (**Figure 11**, bottom), participant-specific peak accuracies still

TABLE 2 | Participant-specific classification results of the water-based recordings.

	Calibration set			Evaluation set		
#	Peak (%)	STD (%)	Time (s)	Peak (%)	STD (%)	Time (s)
V01	68.2	7.9	0.9	69.0	11.1	0.9
V02	74.4	8.2	0.8	72.2	11.3	0.9
V03	60.1	8.9	0.8	56.9	7.7	0.8
V04	59.4	7.4	0.6	64.5	8.3	0.6
V05	60.7	7.3	0.8	58.1	8.5	0.8
V06	76.4	7.4	0.9	75.3	11.1	1.1
V07	68.3	8.2	1.3	66.7	11.1	1.0
V08	74.1	6.8	1.1	55.1	7.7	1.1
V09	59.7	7.5	0.8	63.4	11.5	0.9
V10	62.8	8.3	0.8	62.5	6.9	0.8
V11	68.0	8.0	0.9	64.9	10.3	1.0
V12	60.7	8.9	0.5	55.6	8.2	0.6
V13	60.9	7.6	0.4	60.0	6.9	0.3
V14	65.0	8.6	1.6	57.1	9.0	1.5
V15	61.9	7.7	0.9	53.7	8.5	0.8
Average	65.4	7.9	0.9	62.3	9.2	0.9

Columns 2–4 show peak performance (%), standard deviation (%), and time of occurrence (s) with respect to the movement onset for the calibration set. Columns 5–7 show the same for the test set.

yielded on average 56.4%. **Table 3** depicts participant-specific performance results in detail.

Comprehensive Analysis

Behavioral Analysis

We analyzed the duration of the reach-and-grasp actions (see **Figure 12**). The time information was provided by the instrumentalized objects and extracted from all trials. Then, for each participant and grasp type (palmar, lateral) the average duration was calculated. We were interested in testing, for each grasp type, the possible time differences among the three recording systems. To do so, we computed two separate one-way ANOVAs (for each grasp type) with three levels (gel, water, dry).

A significant effect was found in the lateral grasp [$F(2,42) = 4.6$; $p = 0.016$]. *Post hoc* pairwise multiple comparison tests using the Tukey–Kramer criterion showed a significant effect between the gel- and water-based recordings ($p = 0.015$).

Performance Analysis

We used a one-way ANOVA to compare the differences in classification accuracies among the three EEG recordings. We used separate ANOVAs for the training and test data. We found a significant effect in the training data [$F(2,42) = 5.86$; $p = 0.006$]. *Post hoc* pairwise comparisons revealed a significant performance decrease in the dry-electrodes recordings with respect to the gel- ($p = 0.037$) and water-based recordings ($p = 0.006$). Similarly, we found a significant effect in the test data [$F(2,42) = 4.14$; $p = 0.023$], and *post hoc* pairwise comparisons revealed a significant performance decrease in the dry-electrodes recordings with respect to the gel- (at statistical trend, $p = 0.08$) and water-based ($p = 0.026$) recordings.

Finally, we repeated the previous statistical analysis with the classification accuracies obtained when only using the sensorimotor channels for classification ($n = 11$). The ANOVAs testing showed no significant effects (in either training or test data) among the three EEG recordings. Participant specific performance results can be found in the **Supplementary Tables S1, S2**.

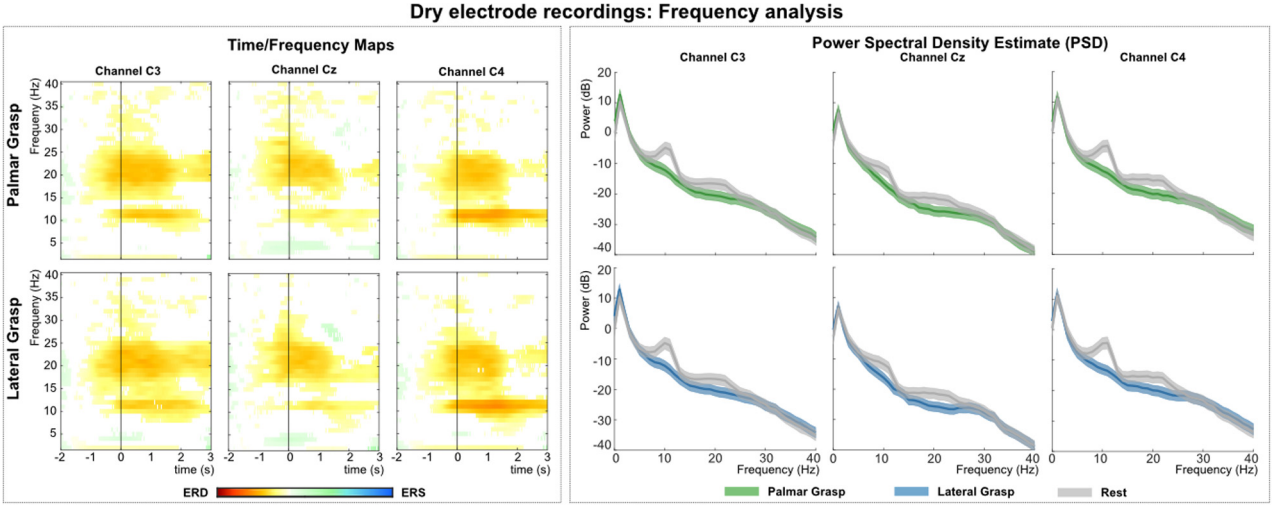


FIGURE 9 | Frequency analysis of the dry-electrodes recordings. **(Left)**: Grand average of the time-frequency maps (Graimann et al., 2002) for movement conditions for positions C3, Cz, and C4 with respect to the reference period (–2 –1) s. The black vertical line represents the movement onset. Hot colors show significant ERD (cold colors represent significant relative increase in power [event-related synchronization (ERS)]. Significant differences with respect to the reference period were calculated using non-parametric t-percentile bootstrap statistics (alpha = 0.05). **(Right)**: Grand average of the PSD calculation of the reaching phase [0 1.5] s. Colored lines represent the PSDs of the movement conditions, gray lines show the PSD of the rest condition. The shaded areas show 95% confidence intervals which were calculated using non-parametric t-percentile bootstrap statistics (alpha = 0.05).

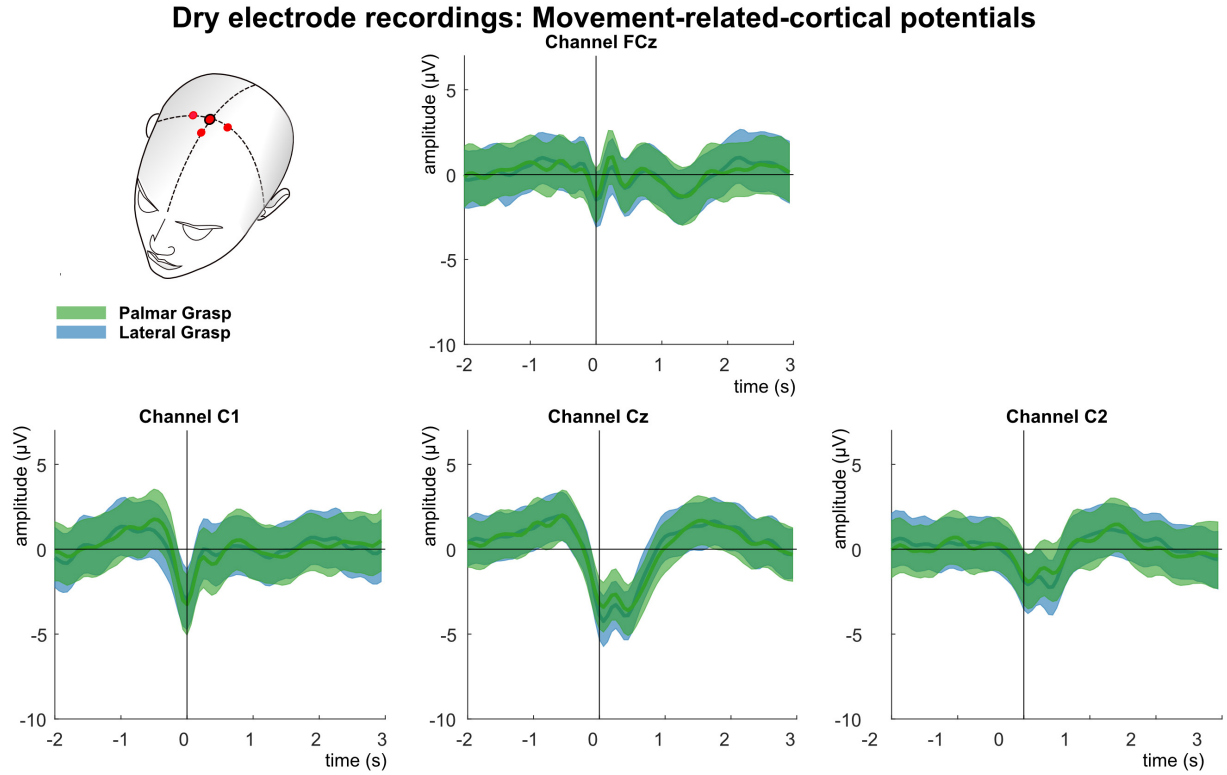


FIGURE 10 | Movement related cortical potentials (MRCPs) of the dry-electrode recordings. Grand averages (bold lines) and 95% confidence intervals (shaded areas) for palmar (green) and lateral (blue) grasp conditions. Channels shown are FCz, C1, Cz, and C2. The black perpendicular line represents the movement onset.

DISCUSSION

This study confirmed that EEG based correlates of reach-and-grasp actions can be successfully identified from the LFTD and time-frequency domain using the water-based EEG-Versatile^{TM} system and the dry-electrodes EEG-Hero^{TM} headset. In addition, we provided a gel-based recording in a laboratory environment (gold standard), which followed the same experimental parameters. In a single-trial multiclass based decoding approach, which incorporated both movement conditions and rest, we could show that the LFTD correlates were also decodable. Grand average peak accuracy calculated on unseen test data yielded for the water-based electrode system 62.3% (9.2% STD), whereas for the dry-electrodes headset reached 56.4% (8% STD). For the gel-based electrode system 61.3% (8.6% STD) could be achieved. The adjusted chance level for this decoding approach was 45.7%, adjusted Wald interval, alpha = 0.05 (Müller-Putz et al., 2008; Breitwieser et al., 2012).

A quantifying comparison between the individual systems is hardly possible due to inter-subject variations and technical factors such as the number of channels for calculating the spatial filters in preprocessing (Gel: 58; Water: 32; Dry: 11) that might influence the outcome. Taking this consideration into account, the gel-based and the water-based system yielded comparable decoding performances and, despite the decreased number of channels of the dry-electrodes headset, the average performance decreased only by less than 6%. Apart from these investigations, we leave it open to the reader to compare systems. For this we provide the complete data sets of all recordings so that the interested community can make their own conclusions.

Time Frequency and PSD Analysis

Calculated ERD/S maps (Graimann et al., 2002) show significant relative power changes for all recording systems. In general, the relative power decrease (ERD) starts already up to 1 s before the actual movement onset and is most prominent in the alpha (mu) and beta frequency bands (Andrew and Pfurtscheller, 1995; Florian and Pfurtscheller, 1995). For the grand average of the gel-based recordings this power decrease is also pronounced stronger, i.e., in terms of frequency range on the contralateral side to the executing right hand than on central or ipsilateral locations. Gel-based recordings showed a more pronounced ERD on the contralateral side in beta frequency bands, which is an expected effect reported in numerous studies (Pfurtscheller and Lopes da Silva, 1999; Pfurtscheller and Neuper, 2001; Müller-Putz et al., 2010). However, this phenomenon was not visible for the water-based and dry electrodes, rather showing a pronounced bilateral ERD (Zaepffel et al., 2013). Moreover, the ERD, especially in the alpha band around 8–12 Hz is pronounced stronger on the ipsilateral side in all three types of recordings (see **Supplementary Figure S1**). The grand average results of the PSD analysis show for all three investigated systems significant

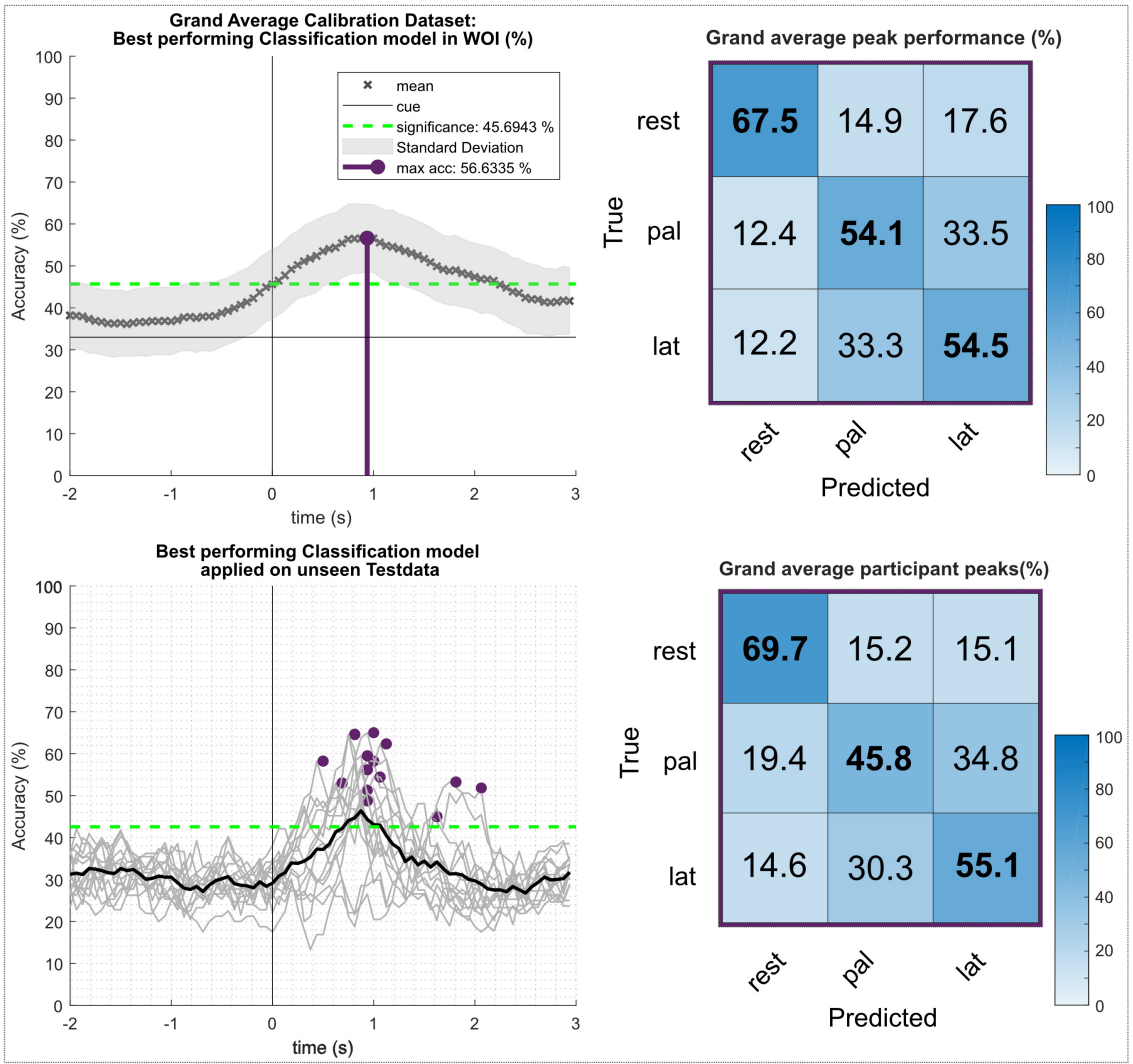


FIGURE 11 | Single trial decoding performance of the dry-electrodes recordings. **(Top left)** Grand average for the calibration data set. Black crosses show the mean performances for the designated time point. On average, best classification performance could be reached 1s after the movement onset (perpendicular purple line). **(Top right)** Row wise normalized confusion matrix for grand average peak performance. **(Bottom left)** Participant-specific classification results (gray lines) and grand average (black bold line) of the best performing classification model applied on the unseen test data set. Purple dots show participant-specific peak performances. **(Bottom right)** Row wise normalized confusion matrix for the individual peak performances.

power decreases for movement conditions when compared to the rest condition. These differences manifest again in the alpha and beta band range and are pronounced strongest on the contralateral side for the gel- and water-based recordings, thus indicating a lateralization effect measured in absolute power. Regarding the dry electrodes, this lateralization phenomenon was not found as a similar power decrease was measured in both bilateral sides.

Movement-Related Cortical Potentials

Analysis of the MRCPs reveal on a grand average basis a strong similarity between the gel-based and the water-based recordings. Around 1 s before the movement onset a negative deflection can be seen, most pronounced over the central motor cortex (Bereitschaftspotential) (Kornhuber and Deecke, 1964; Shibasaki et al., 1980; Shibasaki and Hallett, 2006). This deflection reaches its peak at the movement onset (time = 0 s). It is strongest over central channel Cz and on the contralateral side to the executing right hand. For both systems, a reafferent positive potential around 300 ms after the movement onset can be observed. It has already been found in previous studies concerning reach-and-grasp actions (Schwarz et al., 2018, 2019). Around 1–1.5 s, a second positivity occurs before the potentials return to baseline. In contrast, the MRCPs of the dry electrode recordings are on grand average smaller and their characteristics, such as the

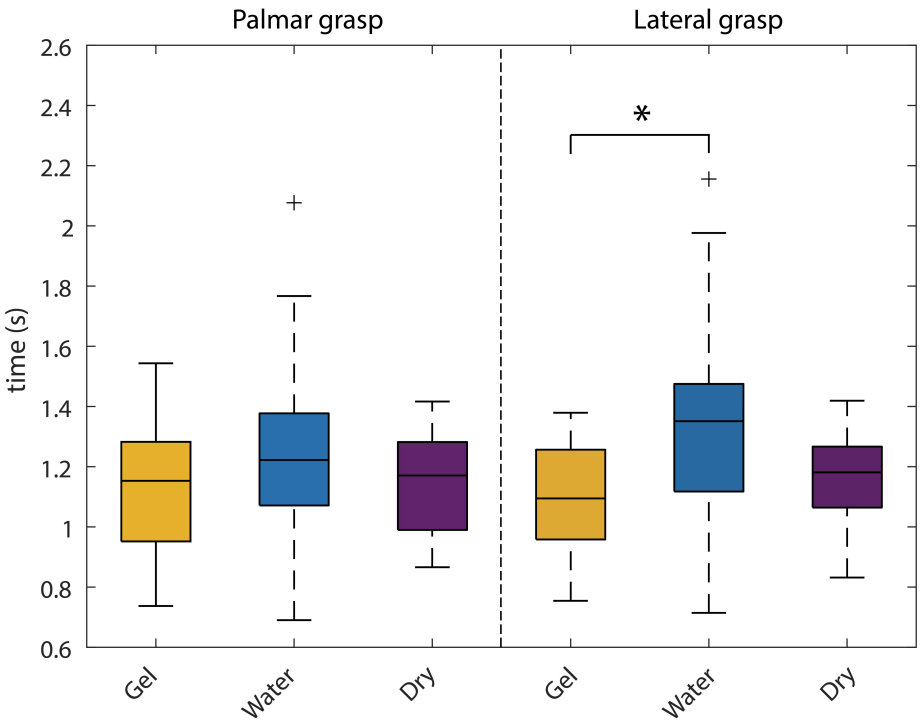


FIGURE 12 | Behavioral analysis. Box-plots of the reach-and-grasp durations for all the systems. The left part shows the palmar grasp durations, whereas the right part shows the lateral grasp ones. Statistical comparisons were carried out for each grasp type separately and significant differences are marked with an asterisk.

TABLE 3 | Participant-specific classification results of the dry-electrodes recordings.

	Calibration set			Evaluation set		
#	Peak (%)	STD (%)	Time (s)	Peak (%)	STD (%)	Time (s)
H01	63.4	8.2	1.9	53.2	7.5	1.8
H02	53.8	9.6	0.8	58.2	8.0	0.4
H03	56.4	8.2	1.1	58.2	9.2	0.9
H04	55.1	8.2	0.9	54.4	7.3	1.0
H05	54.1	8.5	0.8	48.8	6.6	0.9
H06	55.3	7.7	0.9	53.0	7.1	0.6
H07	64.5	7.5	0.6	64.6	8.1	0.8
H08	49.8	8.0	1.6	45.0	4.9	1.6
H09	65.9	8.0	0.9	62.3	7.9	1.1
H10	63.3	7.7	0.9	65.0	11.4	0.9
H11	62.3	8.4	0.7	64.6	7.8	0.8
H12	62.1	6.7	0.8	51.3	8.3	0.9
H13	56.8	6.7	1.8	51.8	8.7	2.1
H14	54.8	9.1	0.5	56.2	8.9	0.9
H15	57.5	8.2	0.7	59.5	8.5	0.9
Average	58.3	8.1	1.0	56.4	8.0	1.0

Columns 2–4 show peak performance (%), standard deviation (%) and time of occurrence (s) with respect to the movement onset for the calibration set. Columns 5–7 show the same for the test set.

BP or the reafferent potential, although clearly identifiable, are attenuated in comparison.

Single-Trial Decoding

The offline classification followed closely the approach initially described in Schwarz et al. (2019) and was primarily designed to simulate a BCI scenario. Using 66% of all available data, we attempted to find the best performing classification model within a WOI. Thereafter, we applied the best performing model on the previously unseen last third of the recorded trials. Regardless of the recording system, all study participants scored significantly higher than the adjusted chance level (calibration set: 45.8%; test set: 42.8%) on both calibration and test. Comprehensive statistical analysis of the participant-specific peak performances showed no significant differences between gel-based and water-based recordings in both sets. Regarding the dry-electrodes headset, a significant decreased performance was found in comparison to the gel- and water-based systems on the calibration set, as well as to the water-based system on the test set.

This decrease in performance for the dry-electrode system was not unexpected, since the number of available electrodes in the dry electrodes (n = 11) is many times smaller than for gel-based (n = 58) or the water-based (n = 32) systems. In a previous study we have already investigated the effect of decreasing the number of (gel based) electrodes available for decoding (Schwarz et al., 2018): We could show that the difference in performance between 61 gel-based electrodes (covering frontal, central, and parietal areas of the scalp) and only 25 gel-based covering sensorimotor areas is minimal. However, further reducing the available electrodes to 15 led to a performance decrease comparable to the dry-electrodes recordings in the current study.

This goes in line with our second classification approach, where we used only the same 11 electrodes in all recording systems, positioned over sensorimotor areas. For both gel-based and water-based electrode systems, the performance was in the same range as for the dry-electrodes headset (see **Supplementary Tables S1, S2**). No significant differences in performance between recording systems could be found anymore. Unfortunately, a direct comparison to other reach-and-grasp studies such as (Agashe et al., 2015; Randazzo et al., 2015; Iturrate et al., 2018) is difficult due significant differences in experimental setup and paradigm and hence cannot be made in a serious manner.

Corresponding Data Sets

The current manuscript is accompanied by a data set of in total 45 study participants, 15 per EEG system. In addition to reproducibility, these datasets will allow analyses beyond the basic analysis steps presented in this manuscript and we encourage the scientific community to try and evaluate new approaches. The datasets are publicly available in the BNCI Horizon 2020 database[2].

Study Limitations

In the current study, we investigated whether EEG-based correlates of reach-and-grasp actions can be successfully identified and decoded from three different EEG systems. However, due to differences in the amount of available EEG and EOG channels, preprocessing and artifact handling could not be performed uniformly. We applied an extended Infomax algorithm on gel-based and water-based recordings and removed ocular based components by visual inspection. This approach could not be performed on the dry electrode recordings, due to the unfavorable number of channels and subsequent number of ICA components, which did not allow a clean separation between ocular and brain activity. Due to the multicentric design of the study, we did not perform the evaluation of the EEG systems on one participant population. Instead, we performed the experiments for each EEG system on an independent group of 15 study participants. Furthermore, the object positions between the gel-based and the water-dry electrodes systems were not exactly replicated and is considered a minor deviation from the original experimental protocol.

CONCLUSION

We presented an EEG dataset on natural reach-and-grasp actions recorded with three different EEG systems – gel-based, water-based and with dry electrodes.

The accompanying study confirmed that reach-and-grasp actions can be successfully identified from MRCPs and time-frequency domain using a water-based EEG-Versatile^{TM} system and a dry electrodes EEG-Hero^{TM} headset. In addition, we provided results from a gel-based recording in a laboratory environment (gold standard), which followed the same experimental parameters.

In a single-trial multiclass based decoding approach, which incorporated both movement conditions and rest, we could show that the MRCPs were also decodable. Although a quantifying comparison between the individual systems is hardly possible due to inter-subject variations and technical factors such as the different number of channels among systems, the gel-based and the water-based system yielded comparable decoding performances. Despite the decreased number of channels of the dry electrodes recordings, the average performance decreased only by less than 6%. Apart from these investigations, we also provide the complete data sets of in total 45 study participants so that the interested community can make their own conclusions. The data set is open access and available at the BNCI Horizon 2020 data base[2].

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: http://bnci-horizon-2020.eu/database/data-sets.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Medical University of Graz (EK: 30-439 ex 17/18). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AS and CE recorded and analyzed the data. AS wrote the original draft. All authors designed the study together and were commited in editing and reviewing the manuscript.

FUNDING

This work was partly supported by EU Horizon 2020 Project MoreGrasp ("643955") and ERC-Cog2015 681231 "Feel Your Reach."

ACKNOWLEDGMENTS

The authors thank Maria Katharina Höller for assistance in measurements as well as Rüdiger Rupp and Katharina Haring for their help on the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2020.00849/full#supplementary-material

[2]http://bnci-horizon-2020.eu/database/data-sets

REFERENCES

Agashe, H. A., Paek, A. Y., Zhang, Y., and Contreras-Vidal, J. L. (2015). Global cortical activity predicts shape of hand during grasping. *Front. Neurosci.* 9:121. doi: 10.3389/fnins.2015.00121

Anderson, K. D. (2004). Targeting recovery: priorities of the spinal cord-injured population. *J. Neurotrauma* 21, 1371–1383. doi: 10.1089/neu.2004.21.1371

Andrew, C., and Pfurtscheller, G. (1995). Event-related coherence during finger movement: a pilot study. *Biomedizinische Technik. Biomed. Eng.* 40, 326–332. doi: 10.1515/bmte.1995.40.11.326

Breitwieser, C., Daly, I., Neuper, C., and MŁuller-Putz, G. R. (2012). "Proposing a standardized protocol for raw biosignal transmission," *IEEE Transactions on Biomedical Engineering*, vol. 59, 852.

Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kübler, A., et al. (1999). A spelling device for the paralysed. *Nature* 398, 297–298. doi: 10.1038/18581

Blankertz, B., Lemm, S., Treder, M., Haufe, S., and Müller, K.-R. (2011). Single-trial analysis and classification of ERP components—a tutorial. *NeuroImage* 56, 814–825. doi: 10.1016/j.neuroimage.2010.06.048

Breitwieser, C., Kreilinger, A., Neuper, C., and Müller-Putz, G. R. (2010). "The TOBI hybrid BCI-the data acquisition module," in *Proceedings of the First TOBI Workshop*, 58. https://www.researchgate.net/profile/Christa_Neuper/publication/267801490_The_TOBI_Hybrid_BCI_-_The_Data_Acquisition_Module/links/55b6720008ae092e9656c81d.pdf

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Di Flumeri, G., Aricò, P., Borghini, G., Sciaraffa, N., Di Florio, A., and Babiloni, F. (2019). The dry revolution: evaluation of three different eeg dry electrode types in terms of signal spectral features, mental states classification and usability. *Sensors* 19:1365. doi: 10.3390/s19061365

Faller, J., Vidaurre, C., Solis-Escalante, T., Neuper, C., and Scherer, R. (2012). Autocalibration and recurrent adaptation: towards a plug and play online ERD-BCI. *IEEE Trans. Neural Syst. Rehabil. Eng.* 20, 313–319. doi: 10.1109/tnsre.2012.2189584

Florian, G., and Pfurtscheller, G. (1995). Dynamic spectral analysis of event-related EEG data. *Electroencephalogr. Clin. Neurophysiol.* 95, 393–396. doi: 10.1016/0013-4694(95)00198-8

Graimann, B., Huggins, J. E., Levine, S. P., and Pfurtscheller, G. (2002). Visualization of significant ERD/ERS patterns in multichannel EEG and ECoG data. *Clin. Neurophysiol.* 113, 43–47. doi: 10.1016/s1388-2457(01)00697-6

Guger, C., Krausz, G., Allison, B. Z., and Edlinger, G. (2012). Comparison of dry and gel based electrodes for P300 brain–computer interfaces. *Front. Neurosci.* 6:60. doi: 10.3389/fnins.2012.00060

Halder, S., Pinegger, A., Käthner, I., Wriessnegger, S. C., Faller, J., Pires Antunes, J. B., et al. (2015). Brain-controlled applications using dynamic P300 speller matrices. *Artif. Intell. Med.* 63, 7–17. doi: 10.1016/j.artmed.2014.12.001

Iturrate, I., Chavarriaga, R., Pereira, M., Zhang, H., Corbet, T., Leeb, R., et al. (2018). Human EEG reveals distinct neural correlates of power and precision grasping types. *NeuroImage* 181, 635–644. doi: 10.1016/j.neuroimage.2018.07.055

Jochumsen, M., Knoche, H., Kidmose, P., Kjær, T. W., and Dinesen, B. I. (2020). Evaluation of EEG headset mounting for brain-computer interface-based stroke rehabilitation by patients, therapists, and relatives. *Front. Hum. Neurosci.* 14:13. doi: 10.3389/fnhum.2020.00013

Jochumsen, M., Knoche, H., Kjaer, T. W., Dinesen, B., and Kidmose, P. (2020). EEG headset evaluation for detection of single-trial movement intention for brain-computer interfaces. *Sensors* 20:2804. doi: 10.3390/s20102804

Jochumsen, M., Niazi, I. K., Dremstrup, K., and Kamavuako, E. N. (2016). Detecting and classifying three different hand movement types through electroencephalography recordings for neurorehabilitation. *Med. Biol. Eng. Comput.* 54, 1491–1501. doi: 10.1007/s11517-015-1421-5

Kaufmann, T., Herweg, A., and Kübler, A. (2014). Toward brain-computer interface based wheelchair control utilizing tactually-evoked event-related potentials. *J. Neuroeng. Rehabil.* 11:7. doi: 10.1186/1743-0003-11-7

Kobler, R. J., Sburlea, A. I, and Müller-Putz, G. R. (2018). Tuning characteristics of low-frequency EEG to positions and velocities in visuomotor and oculomotor tracking tasks. *Sci. Rep.* 8:17713. doi: 10.1038/s41598-018-36326-y

Kornhuber, H. H., and Deecke, L. (1964). Hirnpotentialanderungen beim Menschen vor und nach Willkurbewegungen dargestellt mit Magnetbandspeicherung und Ruckwartsanalyse. *Pflugers Arch. Eur. J. Physiol.* 281:52.

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

Meng, J., Zhang, S., Bekyo, A., Olsoe, J., Baxter, B., and He, B. (2016). Noninvasive electroencephalogram based control of a robotic arm for reach and grasp tasks. *Sci. Rep.* 6:38565.

Millán, J. D. R., Rupp, R., Müller-Putz, G. R., Murray-Smith, R., Giugliemma, C., Tangermann, M., et al. (2010). Combining brain-computer interfaces and assistive technologies: state-of-the-art and challenges. *Front. Neurosci.* 4:161. doi: 10.3389/fnins.2010.00161

Müller-Putz, G. R., Ofner, P., Pereira, J., Pinegger, A., Schwarz, A., Zube, M., et al. (2019). Applying intuitive EEG-controlled grasp neuroprostheses in individuals with spinal cord injury: Preliminary results from the MoreGrasp clinical feasibility study. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2019, 5949–5955.

Müller-Putz, G. R., Scherer, R., Brunner, C., Leeb, R., and Pfurtscheller, G. (2008). Better than Random? A closer look on BCI results. *Int. J. Biomag.* 10:5.

Müller-Putz, G. R., Scherer, R., Pfurtscheller, G., and Neuper, C. (2010). Temporal coding of brain patterns for direct limb control in humans. *Front. Neurosci.* 4:34. doi: 10.3389/fnins.2010.00034

Müller-Putz, G. R., Scherer, R., Pfurtscheller, G., and Rupp, R. (2005). EEG-based neuroprosthesis control: a step towards clinical practice. *Neurosci. Lett.* 382, 169–174. doi: 10.1016/j.neulet.2005.03.021

Müller-Putz, G. R., Schwarz, A., Pereira, J., and Ofner, P. (2016). From classic motor imagery to complex movement intention decoding: the noninvasive Graz-BCI approach. *Prog. Brain Res.* 228, 39–70. doi: 10.1016/bs.pbr.2016.04.017

Ofner, P., Schwarz, A., Pereira, J., and Müller-Putz, G. R. (2017). Upper limb movements can be decoded from the time-domain of low-frequency EEG. *PLoS One* 12:e0182578. doi: 10.1371/journal.pone.0182578

Ofner, P., Schwarz, A., Pereira, J., Wyss, D., Wildburger, R., and Müller-Putz, G. R. (2019). Attempted arm and hand movements can be decoded from low-frequency EEG from persons with spinal cord injury. *Sci. Rep.* 9:7134.

Oostenveld, R., and Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clin. Neurophysiol.* 112, 713–719. doi: 10.1016/s1388-2457(00)00527-7

Pfurtscheller, G., and Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* 110, 1842–1857. doi: 10.1016/s1388-2457(99)00141-8

Pfurtscheller, G., Müller, G. R., Pfurtscheller, J., Gerner, H. J., and Rupp, R. (2003). "Thought"–control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia. *Neurosci. Lett.* 351, 33–36. doi: 10.1016/s0304-3940(03)00947-9

Pfurtscheller, G., and Neuper, C. (2001). Motor imagery and direct brain-computer communication. *Proc. IEEE* 89, 1123–1134. doi: 10.1109/5.939829

Pinegger, A., Faller, J., Halder, S., Wriessnegger, S. C., and Müller-Putz, G. R. (2015). Control or non-control state: that is the question! An asynchronous visual P300-based BCI approach. *J. Neural Eng.* 12:014001. doi: 10.1088/1741-2560/12/1/014001

Pinegger, A., Wriessnegger, S. C., Faller, J., and Müller-Putz, G. R. (2016). Evaluation of different EEG acquisition systems concerning their suitability for building a brain–computer interface: case studies. *Front. Neurosci.* 10:401. doi: 10.3389/fnins.2016.00441

Randazzo, L., Iturrate, I., Chavarriaga, R., Leeb, R., and Del Millan, J. R. (2015). Detecting intention to grasp during reaching movements from EEG. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2015, 1115–1118.

Rohm, M., Schneiders, M., Müller, C., Kreilinger, A., Kaiser, V., Müller-Putz, G. R., et al. (2013). Hybrid brain-computer interfaces and hybrid neuroprostheses for restoration of upper limb functions in individuals with high-level spinal cord injury. *Artif. Intell. Med.* 59, 133–142. doi: 10.1016/j.artmed.2013.07.004

Scherer, R., Billinger, M., Wagner, J., Schwarz, A., Hettich, D. T., Bolinger, E., et al. (2015). Thought-based row-column scanning communication board for individuals with cerebral palsy. *Ann. Phys. Rehabil. Med.* 58, 14–22. doi: 10.1016/j.rehab.2014.11.005

Schwarz, A., Höller, M. K., Pereira, J., Ofner, P., and Müller-Putz, G. R. (2020). Decoding hand movements from human EEG to control a robotic arm in a

simulation environment. *J. Neural Eng.* 17:036010. doi: 10.1088/1741-2552/ab882e

Schwarz, A., Ofner, P., Pereira, J., Sburlea, A. I., and Müller-Putz, G. R. (2018). Decoding natural reach-and-grasp actions from human EEG. *J. Neural Eng.* 15:016005. doi: 10.1088/1741-2552/aa8911

Schwarz, A., Pereira, J., Kobler, R., and Muller-Putz, G. R. (2019). Unimanual and bimanual reach-and-grasp actions can be decoded from human EEG. *IEEE Trans. Bio Med. Eng.* 67, 1684–1695. doi: 10.1109/TBME.2019.2942974

Schwarz, A., Scherer, R., Steyrl, D., Faller, J., and Muller-Putz, G. R. (2015). "A co-adaptive sensory motor rhythms brain-computer interface based on common spatial patterns and Random Forest," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Vol. 2015 (Milan: IEEE), 1049–1052.

Shibasaki, H., Barrett, G., Halliday, E., and Halliday, A. M. (1980). Components of the movement-related cortical potential and their scalp topography. *Electroencephalogr. Clin. Neurophysiol.* 49, 213–226. doi: 10.1016/0013-4694(80)90216-3

Shibasaki, H., and Hallett, M. (2006). What is the bereitschaftspotential? *Clin. Neurophysiol.* 117, 2341–2356. doi: 10.1016/j.clinph.2006.04.025

Skrandies, W. (1990). Global field power and topographic similarity. *Brain Topogr.* 3, 137–141. doi: 10.1007/bf01128870

Snoek, G. J., IJzerman, M. J., Hermens, H. J., Maxwell, D., and Biering-Sorensen, F. (2004). Survey of the needs of patients with spinal cord injury: impact and priority for improvement in hand function in tetraplegics. *Spinal Cord* 42, 526–532. doi: 10.1038/sj.sc.3101638

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 767–791.

Zaepffel, M., Trachel, R., Kilavik, B. E., and Brochier, T. (2013). Modulations of EEG beta power during planning and execution of grasping movements. *PLoS One* 8:e60060. doi: 10.1371/journal.pone.0060060

Conflict of Interest: CE and LM were employed by Bitbrain, Zaragoza, Spain.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 18 September 2020
doi: 10.3389/fnins.2020.568104

BCIAUT-P300: A Multi-Session and Multi-Subject Benchmark Dataset on Autism for P300-Based Brain-Computer-Interfaces

Marco Simões[1,2†], Davide Borra[3*†], Eduardo Santamaría-Vázquez[4,5], GBT-UPM[5,6], Mayra Bittencourt-Villalpando[7], Dominik Krzemiński[8], Aleksandar Miladinović[9], Neural_Engineering_Group[10,11], Thomas Schmid[12], Haifeng Zhao[13], Carlos Amaral[1], Bruno Direito[1], Jorge Henriques[2], Paulo Carvalho[2] and Miguel Castelo-Branco[1*]*

[1] Coimbra Institute for Biomedical Imaging and Translational Research (CIBIT), Institute of Nuclear Sciences Applied to Health (ICNAS), University of Coimbra, Coimbra, Portugal, [2] Centre for Informatics and Systems (CISUC), Department of Informatics Engineering, University of Coimbra, Coimbra, Portugal, [3] Department of Electrical, Electronic and Information Engineering "Guglielmo Marconi" (DEI), University of Bologna, Cesena, Italy, [4] Grupo de Ingeniería Biomédica, Universidad de Valladolid, Valladolid, Spain, [5] Centro de Investigación Biomédica en Red, Biomateriales y Nanomedicina, Madrid, Spain, [6] Biomedical Engineering and Telemedicine Centre, ETSI Telecomunicación, Center for Biomedical Technology, Universidad Politécnica de Madrid, Madrid, Spain, [7] Department of Neurology, University Medical Center Groningen, University of Groningen, Groningen, Netherlands, [8] CUBRIC, School of Psychology, Cardiff University, Cardiff, United Kingdom, [9] Department of Engineering and Architecture, University of Trieste, Trieste, Italy, [10] Department of Biosciences and Bioengineering, Indian Institute of Technology, Guwahati, India, [11] Data Science Research Group, School of Computing, University of Kent, Chatham, United Kingdom, [12] Machine Learning Group, Universität Leipzig, Leipzig, Germany, [13] The University of Sydney, Camperdown, NSW, Australia

There is a lack of multi-session P300 datasets for Brain-Computer Interfaces (BCI). Publicly available datasets are usually limited by small number of participants with few BCI sessions. In this sense, the lack of large, comprehensive datasets with various individuals and multiple sessions has limited advances in the development of more effective data processing and analysis methods for BCI systems. This is particularly evident to explore the feasibility of deep learning methods that require large datasets. Here we present the BCIAUT-P300 dataset, containing 15 autism spectrum disorder individuals undergoing 7 sessions of P300-based BCI joint-attention training, for a total of 105 sessions. The dataset was used for the 2019 IFMBE Scientific Challenge organized during MEDICON 2019 where, in two phases, teams from all over the world tried to achieve the best possible object-detection accuracy based on the P300 signals. This paper presents the characteristics of the dataset and the approaches followed by the 9 finalist teams during the competition. The winner obtained an average accuracy of 92.3% with a convolutional neural network based on EEGNet. The dataset is now publicly released and stands as a benchmark for future P300-based BCI algorithms based on multiple session data.

Keywords: P300, EEG, benchmark dataset, brain-computer interface, autism spectrum disorder, multi-session, multi-subject

Edited by:
Davide Valeriani, Massachusetts Eye and Ear Infirmary and Harvard Medical School, United States

Reviewed by:
Tomasz Maciej Rutkowski, RIKEN Center for Advanced Intelligence Project (AIP), Japan
Motoaki Kawanabe, Advanced Telecommunications Research Institute International (ATR), Japan

***Correspondence:**
Marco Simões
msimoes@dei.uc.pt;
marcoamsimoes@gmail.com
Davide Borra
davide.borra2@unibo.it
Miguel Castelo-Branco
mcbranco@fmed.uc.pt

† These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Neural Technology, a section of the journal Frontiers in Neuroscience

Received: *31 May 2020*
Accepted: *24 August 2020*
Published: *18 September 2020*

Citation:
Simões M, Borra D, Santamaría-Vázquez E, GBT-UPM, Bittencourt-Villalpando M, Krzemiński D, Miladinović A, Neural_Engineering_Group, Schmid T, Zhao H, Amaral C, Direito B, Henriques J, Carvalho P and Castelo-Branco M (2020) BCIAUT-P300: A Multi-Session and Multi-Subject Benchmark Dataset on Autism for P300-Based Brain-Computer-Interfaces. Front. Neurosci. 14:568104. doi: 10.3389/fnins.2020.568104

INTRODUCTION

A Brain-Computer Interface (BCI) is a system that provides a direct communication between the brain and a computer or external device (Wolpaw and Winter Wolpaw, 2012). In short, it must interpret brain activity and translate it into commands that can be used to control devices or programs, from prosthesis, orthosis, wheelchairs and other robots to a mouse or a keyboard in a controlled computer environment (Bamdad et al., 2015; Chaudhary et al., 2016; McFarland and Wolpaw, 2017). Different types of neuroimaging techniques can be used to implement BCIs, i.e., electroencephalography (EEG), magnetoencephalography (MEG), functional Magnetic Resonance Imaging (fMRI), functional Near-Infrared Spectroscopy (fNIRS), among others (Zou et al., 2019). The most common modality is the EEG, since it provides a portable, inexpensive, non-invasive solution to measure brain activity with high temporal resolution (Sitaram et al., 2007; Bhattacharyya et al., 2017; Deshpande et al., 2017; Zou et al., 2019).

There are several approaches to generate brain signals that can be interpreted and transformed into commands by the BCIs, namely event-related potentials (the most prominent being the P300), steady-state visual evoked potentials (SSVEP) or event-related synchronization/desynchronization (ERS/D) through mental imagery. The P300 approach, first attempted by Farwell and Donchin in the 80s (Farwell and Donchin, 1988), uses an oddball paradigm where an infrequent stimulus of interest is presented in a sequence of frequent stimuli of non-interest. With this paradigm, a positive deflection of the EEG measured in the central and posterior parts of the scalp is observed approximately around 300 ms after the infrequent stimulus of interest is presented (Guo et al., 2019; Riggins and Scott, 2019). The most common application of P300-based BCIs is the speller, where a matrix of letters flashing at different times is presented to the user. An infrequent event occurs due to selective attention to a specific target letter. Thus, a P300 potential is elicited whenever the letter the user is paying attention to flashes, and so the target letter can be identified by a P300 detection algorithm and then transmitted. The use-cases of P300-based BCIs have greatly increased over the past years, from steering a wheelchair (Lopes et al., 2016) to composing music (Pinegger et al., 2017).

Despite the wide range of applications, there are still many challenges facing P300-based BCIs to be used more broadly. Achieving portable and practical BCIs that are easy to setup and fast to calibrate is currently a research line of big interest, since it would favorably help the adoption of this new technology in everyday settings (Amaral et al., 2017; Nakanishi et al., 2019; Zou et al., 2019). However, different issues causing low robustness and reliability should be addressed for these systems to be used in real life. Indeed, often low performance is obtained by BCI models, even in laboratory conditions. The noise sensitivity, non-linearity and non-stationarity characteristics of EEG signals represent critical challenges since these properties depend both on the subject and the environment (Yger et al., 2017). As a consequence of non-stationarity, shifts in EEG signals across trials and sessions occur. Therefore, robust feature extraction techniques are needed to overcome these perturbations on the signals (Raza et al., 2019). Moreover, inter-subject variability, due to anatomical and physiological differences among subjects, also represents an important challenge since it hinders the design of participant-agnostic BCIs. Due to these main challenges (intra- and inter-subject variabilities), most BCIs require time-consuming calibrations to maximize their performance, which makes the creation of one-model-fits-all solutions difficult (Saha and Baumert, 2020).

Nevertheless, the methods used for correctly identifying P300 signals have improved in the last years (Lotte et al., 2018). Traditional decoding algorithms rely on separate feature extraction and classification steps. Commonly used P300 features are based on temporal, time-frequency and spatial domains (Demiralp et al., 2001; Bostanov and Kotchoubey, 2006; Agapov et al., 2016), while Linear Discriminant Analysis (LDA), Support Vector Machine (SVM) and Multi-Layer Perceptron (MLP) are the most prominent classifiers used in P300-based BCI approaches. Some examples of recent improvements over traditional methods are the use of Riemannian geometry (Korczowski et al., 2015) or weightless neural networks (Simões et al., 2019). Recently, deep learning techniques were transposed from the computer vision (LeCun et al., 2015) to the EEG decoding field. Among these new solutions, Convolutional Neural Networks (CNN) and CNNs including recurrent layers - such as Long Short-Term Memories (LSTM) - on top of the convolutional extractor were used (CNN-LSTM) (Craik et al., 2019). A key property of these algorithms is that they automatically learn the relevant features for a given task (i.e., the features are learned from the input data without any *a priori* feature extraction and selection) and finalize the target decoding task in an end-to-end fashion (i.e., without separating these steps). Nevertheless, these approaches pose some challenges: they require many hyper-parameters to be tuned (e.g., number of layers, number of kernels, etc.), they introduce a large number of parameters to be optimized during training (which are also difficult to interpret once trained) and thus, require the use of large datasets to achieve state-of-the-art decoding performance (Lawhern et al., 2018; Craik et al., 2019; Zhang et al., 2019). However, few datasets can be found in the literature matching this last requirement.

To evaluate the efficacy of new methods, authors need to compare their results with current state-of-the-art approaches. One viable approach is to implement both their method and established reference methods and apply all of them to the data of interest. Another option is to use benchmark datasets. Benchmark datasets are publicly available data usually launched in competition events where teams have the same information to start with and try to achieve the best possible result with their methods (Rakotomamonjy and Guigue, 2008). These competitions tend to disclose these datasets afterward, allowing both teams and other researchers to continue developing their methods and publish results that are comparable between them, if researchers recreate the original competition conditions on their attempts. Thus, these datasets provide a common ground for the research areas to assess their methods and improve the state-of-the-art.

One important contributor in this field has been the Berlin Brain-Computer Interface (BBCI) group through the organization of BCI competitions[1] (Sajda et al., 2003; Blankertz et al., 2004, 2006; Tangermann et al., 2012). The corresponding datasets have been extensively explored and helped significantly the improvement of methods throughout the years (Lotte et al., 2007, 2018). Nevertheless, those datasets were limited in terms of subjects and sessions-per-subject, thus constraining the development of methods highly dependent on multi-session data.

In the scope of the XV Mediterranean Conference in 2019, the International Federation of Medical and Biological Engineering (IFMBE) launched a scientific competition based on a multi-session dataset of P300-based BCI intervention for young adults with autism spectrum disorder (ASD) (Amaral et al., 2018). This intervention was aimed at the rehabilitation of joint-attention, a core developmental skill that is altered in ASD and impacts other skills like language development (Adamson et al., 2019). Joint-attention refers to the ability of following social attentional cues of other people, so one's attention can be directed by the interlocutor to an external object or event of interest. Amaral et al. (2017) developed an interventional BCI based on P300 signals that uses a virtual environment with a virtual human character and several objects of interest to train the ability of participants to follow the cues of the virtual character to the objects. That system was validated in an interventional pilot study (Amaral et al., 2018) where 15 ASD individuals underwent 7 training sessions with this system. The database resulting from that interventional study supported the 2019 IFMBE scientific challenge and is now made public to the scientific community at https://www.kaggle.com/disbeat/bciaut-p300 (doi: 10.34740/kaggle/dsv/1375326). This paper describes the challenge and corresponding dataset, summarizes the approaches by the competing teams and draws some conclusions from them, challenging the BCI research community to improve the current best performances achieved by the participating teams.

MATERIALS AND METHODS

Experiment Description

Overview of the P300-Based BCI System

The BCI system is composed mainly by two modules: data acquisition module and stimuli presentation module. For the data acquisition module, we used the g.Nautilus system (g.tec medical engineering GmbH, Austria) to record EEG data from 8 active electrodes positioned at C3, Cz, C4, CPz, P3, Pz, P4, POz locations. The reference electrode was placed at the right ear and the ground electrode at AFz location. Sampling rate was set to 250 Hz and data were acquired notch-filtered at 50 Hz and passband-filtered between 2 and 30 Hz. As for the stimuli presentation module, we used the Vizard toolkit to create and display a virtual environment consisting of a bedroom with common type of furniture (shelves, a bed, a table, a chair, and a dresser) and objects (frames, books, lights, a printer, a radio, a ball, a door, a window, and a laptop), as shown in **Figure 1**. The objects used as stimuli throughout the experiment (and their respective labels) were: 1. books on a shelf, 2. a radio on top of a dresser, 3. a printer on a shelf, 4. a laptop on a table, 5. a ball on the ground, 6. a corkboard on the wall, 7. a wooden plane hanging from the ceiling, and 8. a picture on the wall. The virtual environment was presented via the Oculus Rift Development Kit 2 headset (from Oculus VR).

Each block consists of the user trying to identify one of the objects as the target. For that, *K* runs are repeated. One run is composed by a single flash of each object once for 100 ms at different times and random order, with an Inter-Stimulus Interval (ISI) of 200 ms. **Figure 2** provides a schematic for this structure.

BCI Session Flow

Fifteen participants performed 7 identical training sessions in different days, the first four on a weekly basis and the last three on a monthly basis. Each training session was divided in two parts: calibration and online phase. Data from calibration and online phases were named in the dataset as train and test data, respectively.

The calibration phase was composed of 20 blocks, each block containing 10 runs. Because we used 10 runs per block, a total of 200 target P300 signals and 1400 non-target signals were acquired at this phase. With these data, the session-specific classifiers were trained for the online phase and the number of runs per block (K) to use on the online phase was defined. K was defined during the online sessions of the clinical trial as the minimum number of runs for which the classifier achieved an accuracy above 80%, in the calibration data.

Regarding the online phase, 50 blocks were taken for each participant using K runs per block. The value of K varied between subjects and sessions, since it was an output of the calibration phase, ranging from 3 to 10.

Dataset Structure and Contents

The dataset folder structure is organized by subjects, with a folder for each subject named *SBJXX*, with *XX* varying from 01 to 15. Within each subject folder there is a set of folders containing the data from each session, named *SYY*, with *YY* varying from 01 to 07. Each session folder contains a separate folder for the training and testing data, named *Train* and *Test*, respectively. The structure and the contents of train and test folders of each session are described in **Box 1**.

BOX 1 | Dataset Folder Structure.

```
SBJ01\
SBJ02\
..
SBJXX\
        S01\
        S02\
..
        SYY\
                Train\
                Test\
```

[1] http://www.bbci.de/activities#competition

FIGURE 1 | Snapshot of the virtual environment, showing the scenario, the virtual avatar and the objects for joint-attention targets.

Train folder

- **trainData.mat** – Data from the calibration phase, structured as [*channels* x *epoch* x *event*], epoch corresponding to the data samples from −200 ms to 1000 ms relative to the event stimulus onset (epoch length of 1200 ms; 300 data samples).
- **trainEvents.txt** – One label per line (from 1 to 8), corresponding to the order of the flashed objects.
- **trainTargets.txt** – 1 or 0 per line, indicating if the flashed object was the target or not, respectively.
- **trainLabels.txt** – Label of the target object per line (from 1 to 8), one for each block.

Test folder

- **testData.mat** – Data from the online phase, in the same structure as the train data.
- **testEvents.txt** – One label per line (from 1 to 8), corresponding to the order of the flashed objects.
- **testTargets.txt** – 1 or 0 per line, indicating if the flashed object was the target or not, respectively.
- **testLabels.txt** – Label of the target object per line (from 1 to 8), one for each block.
- **runs_per_block.txt** – File containing only one number, corresponding to the number of runs per block used in the online phase (from 3 to 10).

The number of epochs corresponds to # events per run * # runs per block * # blocks. For the training data, it represents 8 events per run * 10 runs per block * 20 blocks = 1600 epochs. As for the test data, since the number of runs varies between sessions, the number of epochs varies in consequence, in a total of 8 events per run * K runs per block * 50 blocks = 400 * K epochs.

The channels' order in the data matrices is C3, Cz, C4, CPz, P3, Pz, P4, POz. The first sample of each epoch corresponds to the

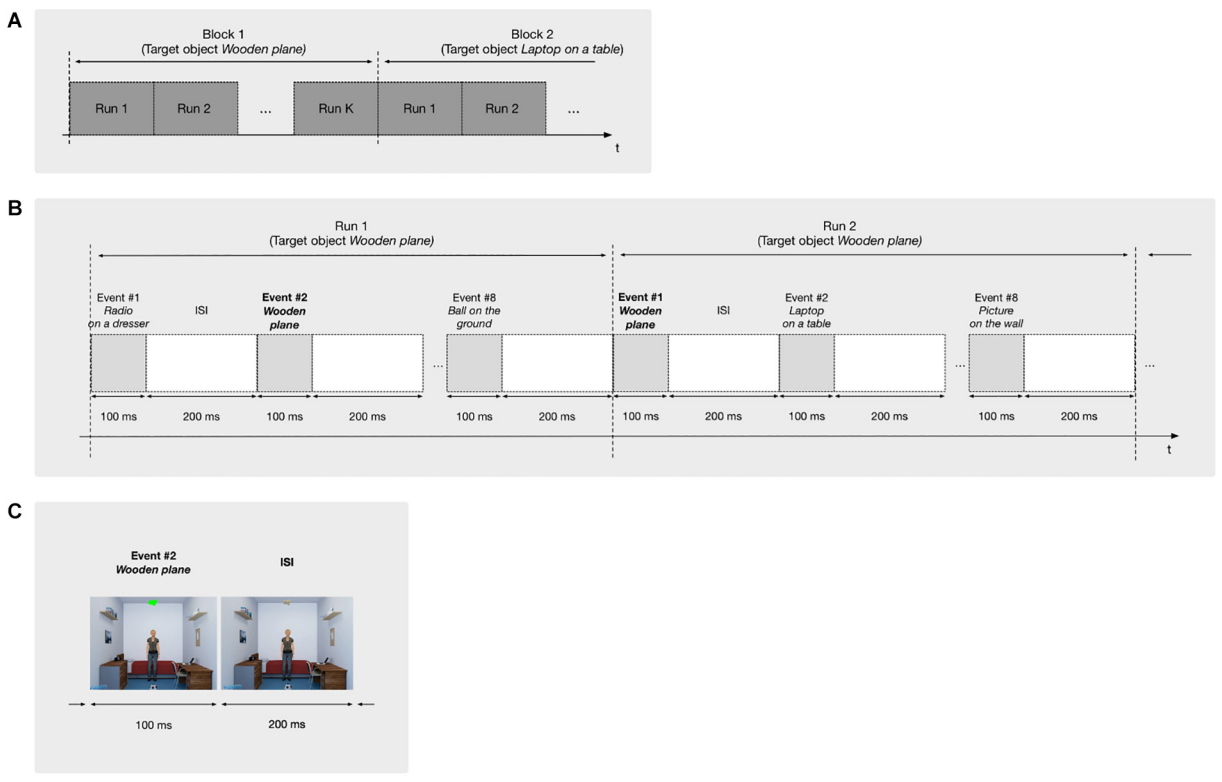


FIGURE 2 | Structure of the paradigm with its subdivisions in blocks, runs and events. **(A)** Structure of the blocks: each block is used to identify a single target object and is composed by K runs. **(B)** Structure of the runs: each run is composed by 8 events, each consisting of the flashing of one of the objects. **(C)** Structure of an event: it consists of the flashing of the corresponding object by 100 ms, followed by an interval of 200 ms.

time −200 ms relative to the stimulus onset and the last sample to corresponds to the time 996 ms after the stimulus onset (the last sample < 1000 ms), with a sampling rate of 250 Hz, for a total of 300 samples.

Challenge Structure

For the 2019 IFMBE Scientific Challenge, teams were asked to maximize the P300-based object detection accuracy for the 7 sessions of the 15 ASD participants of the BCIAUT clinical trial. For each session, a train and test set were created, without disclosing the true labels of the test sets. The challenge was divided into two phases with a different number of attempts per phase (**Table 1**). For *phase I*, sessions 1–3 were provided, without the test labels. At the end of *phase I*, the true test labels of those three sessions were made available to the participants along with the remaining 4 sessions (4–7), the latter without the true test labels (*phase II*). This way, teams could use the true labels of the first three sessions to improve their classifiers, if working with multi-session data. Teams were allowed to submit 5 attempts during phase I and 10 attempts during phase II. The best submission of each team throughout the allowed attempts on each phase was used to rank the teams. The complete dataset (including all true labels) is now available at https://www.kaggle.com/disbeat/bciaut-p300 (doi: 10.34740/kaggle/dsv/1375326).

Submissions and Approaches

Fourteen teams participated in phase I of the competition, while 9 teams participated in phase II and concluded the challenge. The results shown in this manuscript refer to the phase II of the competition. The performance metric used to compare the performance of contesting teams was the target object detection accuracy, computed as the ratio between the number of correct predicted blocks and the total number of blocks to decode. Based on the average target object accuracy across subjects and sessions, the approaches proposed by each team were ranked up.

The following list of IDs reflects the final ranking of the competition:

- ***ID-1***: DB, Silvia Fantozzi and Elisa Magosso (Borra et al., 2020a).
- ***ID-2***: Eduardo Santamaría-Vázquez, Víctor Martínez-Cagigal, Javier Gomez-Pilar and Roberto Hornero (Santamaría-Vázquez et al., 2020).
- ***ID-3***: Lucia de Arancibia, Patricia Sánchez-González, Enrique J. Gómez, M. Elena Hernando and Ignacio Oropesa (de Arancibia et al., 2020).
- ***ID-4***: MB-V and Natasha M. Maurits (Bittencourt-Villalpando and Maurits, 2020).
- ***ID-5***: DK, Sebastian Michelmann, Matthias Treder and Lorena Santamaria (Krzemiński et al., 2020).
- ***ID-6***: AM, Miloš Ajćević, Giulia Silveri, Gaia Ciacchi, Giulietta Morra, Joanna Jarmolowska, Piero Paolo Battaglini and Agostino Accardo (Miladinović et al., 2020).
- ***ID-7***: Bipra Chatterjee, Ramaswamy Palaniappan and Cota Navin Gupta (Chatterjee et al., 2020).
- ***ID-8***: V. Sophie Adama, Schindler Benjamin and TS (Adama et al., 2020).
- ***ID-9***: HZ, Shiduo Yu, Joseph Prinable, Alistair McEwan and Petra Karlsson (Zhao et al., 2020).

TABLE 1 | Timetable and number of attempts for the two phases of the competition.

Phase	Start Date	End Date	Number of Attempts
Phase I	01-03-2019 10:00	15-05-2019 23:59	5
Phase II	20-05-2019 10:00	30-06-2019 23:59	10

For each team, a brief description of the proposed methodology is reported:

- ***ID-1***: Epochs were extracted between −100–1000 ms, and the signals were downsampled to 128 Hz. The decoding solution was based on a CNN performing classification at the level of single trial (EEG response to a single stimulus, without averaging). The input was a 2-D representation composed by the EEG channels along one dimension (spatial dimension) and time steps along the other dimension (temporal dimension). The CNN was an adaptation of EEGNet (Lawhern et al., 2018) trained to discriminate between P300 and non-P300 classes. In this CNN design, depthwise and pointwise convolutions are used to keep the number of trainable parameters limited. The architecture in its fundamental subnetworks and main connections between neurons is displayed in **Figure 3**. Furthermore, a detailed description of these subnetworks including the main hyper-parameters, output activation shapes and number of trainable parameters introduced is reported in **Table 2**. The CNN is composed by 3 main subnetworks (here labeled as A, B, C), performing different operations on the input. These include a temporal and spatial feature extractor (**Figure 3A**) that learns meaningful temporal and spatial filters, a summary feature extractor (**Figure 3B**) that learns to extract temporal summaries for each feature map of the subnetwork A individually; and a classification module (**Figure 3C**) that finalizes the classification task based on the output of the subnetwork B. The obtained single-trial probabilities were then averaged together across runs related to a specific object belonging to each block, and then the object with maximum average probability was selected, solving the target 8-way classification task. Different intra-subject training strategies were explored, including inter-session (i.e., training subject-specific classifiers) and intra-session (i.e., training session-specific classifiers) training strategies. The top-performing solution of *ID-1* was the one adopting a subject-wise inter-session strategy. The code of the CNN and the weights of the trained models are available at https://github.com/ddavidebb/IFMBE2019Challenge-BCIAUT-P300.
- ***ID-2***: EEG signals were epoched between 0–1000 ms, applying a baseline (−200-0 ms) normalization. The input representation is the same as in ***ID-1***. The task was

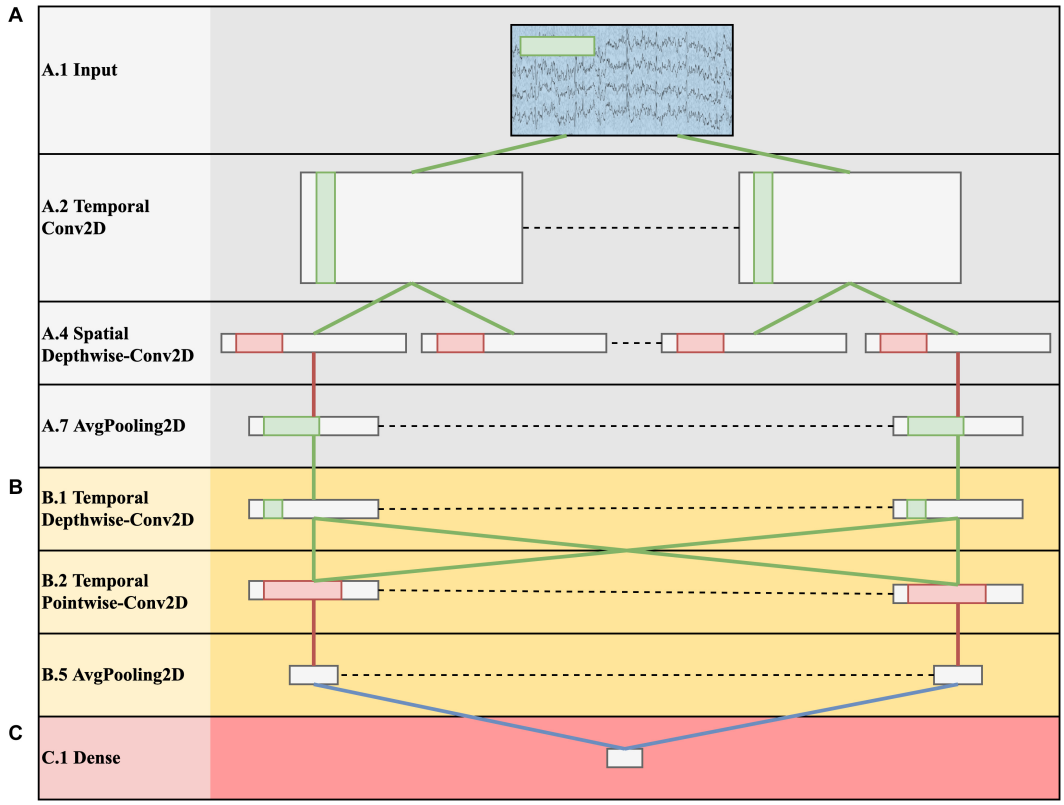


FIGURE 3 | Architecture schematization of the winning solution *ID-1* based on *EEGNet*. The represented shapes correspond to the output of each layer. Green lines represent convolutional connections, red lines pooling connections, and blue lines dense connections. The CNN is composed by a temporal and spatial feature extractor **(A)**, a summary feature extractor **(B)** and a classification module **(C)**.

TABLE 2 | Architecture design inspired from EEGNet and adopted in *ID-1*.

Subnet.	Layer ID	Layer	Hyper-parameters	# pars	Output shape	Activation
A	A.1	Input		0	(1,8,140)	
	A.2	Temporal Conv2D	K = 8, F = (1,65), *P* = (0,32)	520	(8,8,140)	Linear
	A.3	BatchNorm2D		16	(8,8,140)	
	A.4	Spatial Depthwise-Conv2D*	D = 2, K = 16, F = (8,1), *P* = (0,0)	128	(16,1,140)	Linear
	A.5	BatchNorm2D		32	(16,1,140)	
	A.6	Activation		0	(16,1,140)	Exponential Linear Units (ELU)
	A.7	AvgPooling2D	F = (1,4)	0	(16,1,35)	
	A.8	Dropout	*p* = 0.25	0	(16,1,35)	
B	B.1	Temporal Depthwise-Conv2D	D = 1, K = 16, F = (1,17), *P* = (0,8)	272	(16,1,35)	Linear
	B.2	Temporal Pointwise-Conv2D	K = 16, F = (1,1), *P* = (0,0)	256	(16,1,35)	Linear
	B.3	BatchNorm2D		32	(16,1,35)	
	B.4	Activation		0	(16,1,35)	ELU
	B.5	AvgPooling2D		0	(16,1,4)	
	B.6	Dropout	*p* = 0.25	0	(16,1,4)	
C	C.1	Dense	*N* = 2	130	(2)	Linear
	C.2	Activation		0	(2)	Softmax

*K and F are the number and the size of the kernels, respectively. P is the padding size, D the depth multiplier, N the number of neurons in the dense layer and finally p the dropout rate. Light-gray denote layers with trainable parameters. The total number of trainable parameters is 1386. *Unitary kernel max-norm constraint.*

faced as a 2-way classification decoding P300 and non-P300 classes for each trial adopting an adaptation of the CNN proposed by Manor et al. (Manor and Geva, 2015), a CNN-LSTM and a CNN-BLSTM. Furthermore, these deep learning architectures were compared with a more traditional machine learning pipeline including

SWLDA. The top-performing algorithm proposed by ***ID-2*** was CNN-BLSTM. This network was composed of one convolutional layer 1-D that extracts spatio-temporal patterns on the input, two bidirectional LSTM layers and one dense layer. The single-trial probabilities were averaged to obtain object-level probabilities as in ***ID-1***. An intra-subject and inter-session training strategy was adopted, training subject-specific classifiers. The code of the models and the weights of the trained models are available at https://github.com/esantamariavazquez/IFMBE2019Challenge-BCIAUT-P300.

- ***ID-3***: EEG signals related to a specific object were averaged across trials of the same block. Feature extraction was based on temporal and time-frequency parameters. Temporal features were extracted in epochs between 0–1000 ms by downsampling the signals with a decimation factor of 10. In addition to temporal features, features based on continuous wavelet transform (CWT) were extracted from epochs between 200–712 ms. The t-CWT was computed based on a Mexican Hat wavelet on scales corresponding to the delta (0.5–4 Hz) and theta (4–8 Hz) bands (Demiralp et al., 2001; Bostanov and Kotchoubey, 2006). These temporal and time-scale features were concatenated across channels in a single vector. Principal component analysis (PCA) was applied for feature dimensionality reduction, which resulted in a final vector of 120 features. A comparison of different combinations of linear and non-linear machine learning approaches was performed. More specifically, linear discriminant analysis (LDA) and support vector machines with linear kernel (LSVM), and a more complex support vector machine with radial kernel (RSVM) were employed. The object whose corresponding signals yielded a higher probability of containing a P300 event was chosen as predicted target object of the block. In addition, the effect on the accuracy of the number of EEG events averaged was studied. An inter-session training strategy was adopted, comparing both subject-specific and inter-subject classifiers, as well as the use of oversampling and boosting techniques to account for class imbalance. LDA outperformed the other classifiers and was used to classify the target object. Best results were obtained for > 3 events averaged. Training subject-specific classifiers yielded the best performance. Oversampling and boosting did not improve the final performance of the classifiers. The developed code and trained models are available at: http://dev.gbt.tfo.upm.es/ioropesa/ifmbe-scientific-challenge-competition---detection-of-p300/tree/master.
- ***ID-4:*** The approach consisted of the adaptation and parameter optimization of an SVM-based algorithm that was previously developed for a 4-choice BCI (Bittencourt-Villalpando and Maurits, 2018) for target identification. During the first phase of the challenge, the original algorithm was adapted for 8 choices and the pre-processing parameters were defined as follows. First, temporal features were extracted in epochs between 0–1000 ms following each event onset and all channels were concatenated in a single feature vector per event for each participant and session. Then, feature vectors containing EEG signals from target events were pseudo-randomly averaged across blocks belonging to the same session for noise reduction. During the second phase of the challenge, an intra-subject and intra-session training strategy was developed, augmenting the dataset with other sessions' signals, and artificially increasing the number of targets per session by adapting the pseudorandom averaging procedure. Eight parameters related to data augmentation and SVM input parameters were optimized throughout the 9 initial attempts and then compared in terms of accuracy. The parameters' description and settings per attempt are detailed in Bittencourt-Villalpando and Maurits (2020). In the last attempt, the best performing parameter setting was selected, resulting in a customized solution per participant and per session.
- ***ID-5:*** This solution exploited Riemannian framework for EEG signal decoding (Korczowski et al., 2015). The approach was computationally efficient and recently outperformed other common state-of-the-art approaches (Barachant et al., 2010). The Riemannian framework was combined with the ensemble learning. The idea was to build upon many "weak" (under-performing) classifiers and then combine their outcomes to improve the performance of the final model. The ensemble of 8 different data features was constructed by combining 2 different band-pass filters (1–20 Hz or 1–8 Hz), 2 trial lengths (from −200 to 1000 ms or from 0 to 600 ms) and 3 different subsets of electrodes (all, central or posterior only electrodes). Then, the ERP prototypes were created by calculating the ERP for each channel. Next, the regularized covariance matrices of a single trial concatenated with the prototype were computed and the resultant matrices were projected into the tangent space of a reference matrix. Fisher Geodesic Discriminant Analysis (FGDA) was used to project the matrices to a lower-dimensional discriminative subspace. The resultant projections were flattened to vectors and used as the features to the ensemble learning algorithm comprising 400 LDA classifiers. The output probability was aggregated across trials belonging to each object to decode the target per each block. An intra-subject and intra-session training strategy was adopted. The developed code is available at https://github.com/dokato/bci-challange.
- ***ID-6:*** The windows mean approach was used to obtain the temporal features on each trial. These were computed for each electrode on 50 ms windows without overlap from 100–1000 ms. Bayesian logistic regression with automatic relevance determination (VB-ARD) (Drugowitsch, 2013) was used to classify the P300 event on each trial. The method has an advantage over other regularization techniques which need a separate validation set to eliminate irrelevant features. Besides, this approach generates a posterior distribution enabling the authors to model the varying-intercept sparse feature model. The modeling applied in this approach is similar to the one proposed by Bishop (2006) with a variation of Automatic Relevance Determination (ARD) that instead of using type-II maximum likelihood (MacKay, 1992), applies full Bayesian

treatment (Drugowitsch, 2013). The primary generative model matches the one employed in Bishop (2006), and the prior is selected to be non-informative, modeled by a conjugate Gamma distribution (Drugowitsch, 2013). This makes the model parameter-free and easy to use without deep knowledge in the data science domain. The advantage of this methodology is that obtained distribution allows the authors to find the inverse of the predictors' covariance matrix (precision matrix) and apply Automatic Relevance Determination (ARD) that assigns an individual hyper-prior to each regression coefficient separately determining their relevance and produces for each trial a class-belonging probability. Lastly, single-trial probabilities were averaged together across trials for each object belonging and the one with maximum average probability was selected. In this method, an intra-subject and intra-session training strategy was performed. The demo code is available at https://github.com/miladinovic/BCILabTS under subfolder *userscripts*.

- ***ID-7:*** Whole signals were used (−200–1000 ms) and the pre-stimulus mean (−200-0 ms) was removed. Signals were filtered between 2–12 Hz and the filtered signals were downsampled 10-times. Then, these downsampled electrode signals were normalized epoch-wise in the range −100–1000 ms. These temporal features were used to classify the P300 event for each trial with BLDA, RUSBoost and CNN. The best performing classifier for each subject was used (subject-specific classifier). Then, a majority voting was done to determine the target object within each specific block. An intra-subject and inter-session training strategy was performed.
- ***ID-8:*** EEG signals were averaged across trials related to a specific object belonging to each block. Temporal features were extracted for each electrode by averaging for each time window from 200–450 ms and decimating the output with a factor of 12. In addition, Pearson's correlation coefficient was computed for each electrode between the time window of interest and the time window preceding stimulus presentation (−200-0 ms). These temporal features and correlation coefficients were concatenated across channels in a single feature vector. An inter-subject and inter-session training strategy was performed, by which a variety of competing supervised learning techniques (decision tree, random forest, SVM, MLP) were trained to classify the target object within each block. From those, MLP performed best on the given data.
- ***ID-9:*** Epochs were extracted from 0–600 ms. An additional 20 Hz low-pass filter was applied to the original data. In addition, a custom filter was designed to address each subject- and session-specific noise features. The temporal features were selected using a linear support vector regression as a pre-selector for features in the data. A comparison between linear and non-linear methods was performed, using SVM, LDA, 1D 4-layer CNN, 1-layer LSTM. LDA was the top-performing classification algorithm for *ID-9* and was used to classify the P300 event for each trial. Then, the label that appeared most times within each block was the target object to decode. An intra-subject and intra-session training strategy was adopted. The code is available at https://github.com/hyphenzhao/MEDICON2019ScientificChallenge.

A summary of the top-performing method of each team adopted for the challenge is shown in **Table 3**.

Statistical Analysis

For each team, the best-performing solution proposed among the phase II attempts – in terms of target object accuracy averaged across subjects and sessions – was selected for analysis and the algorithms were then ranked up based on this average score. Furthermore, the metrics scored by algorithms *ID-2:9* were compared with the winning algorithm (*ID-1*) using Wilcoxon signed-rank tests. To correct for multiple tests, a false discovery rate correction at 5% using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995) was applied and the corrected *p*-values are reported.

RESULTS

In **Tables 4**, **5** the accuracies of the proposed approaches are shown, describing the decoding variability across subjects and recording sessions. In particular, **Table 4** reports for each subject the average target object accuracy across sessions (i.e., performance at the level of single subjects), while **Table 5** reports for each session the average target object accuracy across subjects (i.e., performance at the level of single session).

Averaging across sessions and across subjects, ***ID-1*** significantly outperformed the other approaches, with less variability across subjects and sessions. Looking at the performance at the level of subjects, ***ID-1*** provided the best performance metric for 14 out of 15 subjects (for subject #4, ***ID-2*** provided a top-performance across the proposed solutions too), while ***ID-3*** provided the best performance metric for 1 out of 15 subjects (subject #14).

Averaging across subjects, ***ID-1*** significantly outperformed the other approaches within each recording session, with less variability across subjects and providing an average performance above 90% for all the phase II sessions.

DISCUSSION

In this study, a large multi-session and multi-subject dataset acquired during a P300-based BCI intervention for young adults with ASD was presented. The evolution and the practical application of deep learning solutions for EEG decoding depend on the availability of large multi-subject datasets. Furthermore, the lack of multi-session datasets hinders the design of reliable algorithms across recording sessions. Thus, the described dataset represents a multi-session collection of signals that can be used as a benchmark to design accurate and reliable data-hungry algorithms, such as deep learning solutions, for P300 decoding tasks.

TABLE 3 | Summary of the best-performing algorithm of each team developed for the challenge.

ID #	acc. (%)	Pre-processing	Methodology	Post-processing	Training strategy	Framework
ID-1	92.3 ± 1.8	• Epochs from -100 to 1000 ms • Downsampling to128 Hz	• CNN based on EEGNet (Lawhern et al., 2018)	• Average probability across runs within a specific block • Decoding of the target object as the object with maximum average probability	• Intra-subject and inter-session	• Python with PyTorch
ID-2	84.3 ± 3.2	• Epochs from 0 to 1000 ms • Baseline normalization from -200 to 0 ms	• CNN-BLSTM	• Average probability across runs within a specific block • Decoding of the target object as the object with maximum average probability	• Intra-subject and inter-session	• Python with Scikit-learn and Keras
ID-3	82.0 ± 2.5	• Temporal features: ◦ Ensemble averaging per block ◦ Temporal epoching from 0 to 1000 ms. ◦ Moving-average downsampling • CWT features: ◦ Temporal epoching from 200 to 712 ms ◦ Most differential points computed with t-Student (t-CWT)	• Temporal features concat (200 features) • Computation of the t-CWT (Bostanov and Kotchoubey, 2006) based on Mexican Hat wavelet (128 points per channel) and CWT features concat. (1024 features) • Feature reduction based onyh PCA (120 features) • LDA	• The object whose corresponding signals yield a higher probability of containing a P300 was chosen as predicted target object of the block	• Intra-subject and inter-session	• MATLAB with Statistics and Machine Learning Toolbox and Signal Processing Toolbox
ID-4	81.5 ± 2.6	• Epochs from 0 to 1000 ms • Pseudorandom • averaging of ERP segments.	• Feature vector with 2000 elements per ERP (concat. of 8 channels*250 elements) • SVM	• The feature vectors were sorted according to the event (flashed object, from 1 to 8) • All runs per block were averaged, per event • The predicted target corresponds to the event with the highest score.	• Intra-subject and intra-session • Data augmentation with other sessions' signals and with pseudorandom averaging	• MATLAB with Statistics and Machine Learning Toolbox 2017.
ID-5	81.2 ± 2.1	• Band-pass filtering with two different filters (1–20 Hz or 1–8 Hz) and two variations of trial length (whole signal or the first 600 ms after stimuli onset) • Three subsets of electrodes were chosen (all, central or posterior electrodes)	• ERP prototypes were created by calculating the ERP for each channel • Regularized covariance matrices of a single trial signal concatenated with prototype were calculated • The resultant covariance matrices were projected into the tangent space of a reference matrix • FGDA was used to project the matrices in tangent space to a lower-dimensional discriminative subspace. These were used as features. • Ensemble of 400 LDA classifiers (taking 40% of data samples and 60% of features) operated on ensemble of signal preprocessed in 8 different combinations	• Aggregated probability of trial belonging to each of the classes.	• Intra-subject and intra-session	• MATLAB

(Continued)

TABLE 3 | Continued

ID #	acc. (%)	Pre-processing	Methodology	Post-processing	Training strategy	Framework
ID-6	80.3 ± 2.2	• Epochs from 100 to 1000 ms	• Temporal features computed on 50 ms windows, without overlap, producing 18 features per channel for each event • VB-ARD	• Average probability across runs within a specific block • Decoding of the target object as the object with maximum average probability	• Intra-subject and intra-session	• MATLAB • BCILAB
ID-7	76.3 ± 2.9	• Epochs from −200–1000 ms • Pre-stimulus mean (−200-0 ms) was removed. • Band-pass filtering 2–12 Hz • Normalization epochwise to the interval [−1,1]	• Temporal features were extracted by downsampling with a factor of 10 the normalized and filtered signals • Three classifiers were trained and tested: ○ BLDA ○ RUSBoost ○ CNN	• The best performing classifier for each subject was used • Majority voting within each run to determine which flash has been classified as target maximum number of time and that was predicted as target for that particular run	• Intra-subject and inter-session	• MATLAB with Classification App RUSBoosted Trees
ID-8	70.0 ± 3.8	• Averaging of EEG signals across trials related to a specific object within each block	• Temporal features [based on (Krusienski et al., 2006)]: averaging within windows from 200–450 ms; 56 features per channel (448 total) • Pearson's correlation coefficients: coefficients were computed between the time window of interest and the time window preceding stimulus presentation (-200-0 ms); 8 features per channel (64 total) • Concatenation of temporal and Pearson's coefficients across channels in a single feature vector MLP	• -	• Inter-subject and inter-session	• MATLAB (pre-processing) • Python with Scikit-learn (main algorithm)
ID-9	67.2 ± 3.3	• Epochs from 0–600 ms • Low-pass filter 20 Hz • Custom filter to address each subject- and session-specific noise features deduced from non-target epochs	• Linear support vector regression as feature pre-selector • LDA	• The label that appeared most times within each block was the target object to decode	• Intra-subject and intra-session	• Python with Scikit-learn

TABLE 4 | Performance at the level of single subject as represented by the average target object accuracies of the best approach proposed by each team.

ID #	Accuracy at the level of single subject (%)																
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**acc (mean ± SEM)**	***p*-value**
ID-1	81	100	86	96	93.5	96	96.5	100	90.5	98	94	84.5	86.5	81.5	100	92.3 ± 1.8	–
ID-2	56	98	67.5	96	80	88	86.5	99	82	93	87.5	80	81	71.5	98.5	84.3 ± 3.2	0.0010
ID-3	73	95	71	91	82.5	86	85	91.5	68.5	88.5	86	80.5	60	84	87.5	82.0 ± 2.5	0.0009
ID-4	64.5	92	68	94.5	84	86	81.5	94	71	87	87	82	66	77	88	81.5 ± 2.6	0.0009
ID-5	69	91	67	88.5	79.5	82.5	83	95	82.5	81.5	85.5	79	69	78	87.5	81.2 ± 2.1	0.0009
ID-6	68	91.5	71.5	92.5	80	84	79	94.5	73.5	82.5	84	78.5	68.5	71.5	85.5	80.3 ± 2.2	0.0009
ID-7	54	93	62.5	90	73	85.5	76	88.5	71	78	80.5	78	65	65	84.5	76.3 ± 2.9	0.0009
ID-8	48	84	58	69	69.5	52	84	94	72	87.5	77	64	50	56.5	84.5	70.0 ± 3.8	0.0009
ID-9	46	85	53	77	65	66.5	67.5	89	57	72	73.5	73	59.5	47	77.5	67.2 ± 3.3	0.0009

The best decoding performance for each subject is colored with light-gray. The mean accuracy (acc) and its standard error (SEM) are reported. Wilcoxon signed-rank test was used to compare ID-1 with ID-2:9 and the corrected p-values for multiple tests are reported.

TABLE 5 | Performance at the level of single session as represented by average target object accuracies across subjects of the best approach proposed by each team.

ID #	Accuracy at the level of single session (%)							
	4		**5**		**6**		**7**	
	acc (mean ± SEM)	***p*-value**	**acc (mean ± SEM)**	***p*-value**	**acc (mean ± SEM)**	***p*-value**	**acc (mean ± SEM)**	***p*-value**
ID-1	92.8 ± 2.4	–	90.4 ± 3.5	–	94.8 ± 1.8	–	91.1 ± 3.0	–
ID-2	85.1 ± 3.1	0.0044	82.0 ± 5.5	0.0026	90.5 ± 2.6	0.0082	79.6 ± 5.6	0.0026
ID-3	81.5 ± 3.3	0.0023	82.0 ± 4.4	0.0062	84.3 ± 2.6	0.0025	80.3 ± 3.7	0.0015
ID-4	80.3 ± 3.0	0.0015	80.7 ± 4.4	0.0037	84.9 ± 2.6	0.0015	80.1 ± 4.2	0.0020
ID-5	79.9 ± 3.3	0.0013	78.4 ± 4.2	0.0015	85.1 ± 2.4	0.0013	81.6 ± 4.2	0.0032
ID-6	78.1 ± 3.6	0.0015	79.6 ± 4.0	0.0017	83.6 ± 2.6	0.0013	80.0 ± 3.7	0.0013
ID-7	75.2 ± 3.8	0.0013	72.8 ± 4.9	0.0013	80.3 ± 2.6	0.0013	76.9 ± 3.6	0.0013
ID-8	70.5 ± 4.3	0.0013	70.3 ± 6.0	0.0013	72.7 ± 3.8	0.0013	66.5 ± 5.7	0.0013
ID-9	64.8 ± 4.3	0.0013	66.9 ± 4.5	0.0013	69.3 ± 4.2	0.0013	67.9 ± 5.4	0.0013

The best decoding performance for each recording session is colored with light-gray. The mean accuracy (acc) and its standard error (SEM) are reported. Wilcoxon signed-rank test was used to compare ID-1 with ID-2:9 and the corrected p-values for multiple tests are reported.

In fact, the richness of the dataset enabled the use of deep learning approaches in the context of the competition. Among the proposed algorithms, a deep learning solution based on a lightweight CNN (see ***ID-1*** in Section "Submissions and Approaches") outperformed both a CNN-BLSTM ($p = 0.001$, across subjects and sessions, see **Table 4**, ***ID-2***) and more traditional machine-learning solutions ($p < 0.001$, across subjects and sessions, see **Table 4**). Furthermore, this was found also for single session recordings ($p < 0.005$ when comparing ***ID-1*** with other solutions, see **Table 5**), with average metrics above 90% (far above the chance level of 12.5%). The best non-deep learning solution adopted temporal and CWT features, alongside with PCA for dimensionality reduction and LDA for classification (see ***ID-3*** in Section "Submissions and Approaches"). The training strategies performed in the approaches ***ID-1***:***3*** were both intra-subject and inter-session. In particular for the winning solution, from the experiments between inter-session and intra-session trainings performed by ***ID-1***, better results were found using all the session signals during the optimization.

When using deep learning approaches with EEG signals, the input representation and the design of spatio-temporal convolutions is not trivial and need to be addressed. Regarding the input representation, the time series are related to electrodes placed on a 3D surface. Typically, EEG signals can be represented in three different ways to feed the input layer of a neural network (Lawhern et al., 2018):

a. Using the original representation of all the available electrode signals to design a 2D representation where EEG channels are reported along one dimension (spatial dimension) and time steps along the other dimension (temporal dimension).
b. Using a transformed representation (e.g., time-frequency decomposition) of all the available electrodes.
c. Using a representation as in (b) with a subset of electrodes.

Among these representations, the first one is preferred since a representation like (b) generally increases the dimensionality (Lawhern et al., 2018), leading to more trainable parameters and, thus, to the need of more data or an increased regularization. Furthermore, several hyper-parameters are introduced depending on the transformation applied. Lastly, representations like (c) share the main disadvantages of (b)

with an additional needing of *a priori* knowledge about the more relevant subset of electrodes to choose. Therefore, representations that respect the scheme (a) are a good compromise between input dimensionality and capability to learn more general EEG features on all the electrode signals (Lawhern et al., 2018). Among the best-performing solutions in this competition, ***ID-1*** and ***ID-2*** adopted the first input representation scheme.

Regarding the design of spatio-temporal convolutions, depending on the information processing in the convolutional module, three different solutions can be designed starting from the input layer:

i. The temporal filtering is performed at first and then the spatial filtering.
ii. The spatial filtering is performed at first and then the temporal filtering.
iii. Mixed spatio-temporal filtering.

The CNN adopted by *ID-1* used the convolutions ordering as in (i), while the CNN-BLSTM adopted by *ID-2* as in (iii). Furthermore, among the solutions proposed by *ID-2*, there was a CNN based on Manor and Geva (2015) adopting a convolution ordering as in (ii). Thus, in this competition, the solutions based on convolution ordering as in (i) outperformed the solutions following (ii) and (iii) designs.

In addition, the layers of the neural network need to be carefully designed to keep control the number of trainable parameters and thus, to avoid overfitting when handling a limited collection of training signals. To this aim, architectures like EEGNet (Lawhern et al., 2018) were proposed including optimized convolutions, such as depthwise and separable convolutions (Chollet, 2016). The CNN adapted in ***ID-1*** was inspired from Lawhern et al. (2018) and introduced only 1386 trainable parameters, while the CNN-BLSTM designed by ***ID-2*** introduced 10113 parameters. Lastly, among the solutions proposed by ***ID-2*** (different from the best-performing algorithm of ***ID-2***), a CNN based on Manor et al. (Manor and Geva, 2015) introduced 37428963 parameters. Therefore, in this competition the use of a lightweight architecture to solve the target P300 decoding task was beneficial. This result is in line with the recent growth of interest in the design of optimized layers in CNNs for EEG decoding as proposed by Zhao et al. (2019) and Borra et al. (2020b).

The **BCIAUT-P300** dataset presents rare characteristics which reinforce its potentialities to work as a benchmark for P300-based BCI methods: 1) the multi-subject dimension, with 15 participants undergoing the same procedure, enable the possibility of developing inter-subject methods for generalized off-the-shelf applications; 2) the multi-session dimension, since each subject repeated the same training task 7 times in different weeks, enables the study of stability and reliability of subject-specific BCI methods throughout time, and even the inclusion of reinforcement learning strategies by approaching the sessions gradually; and 3) the ASD clinical dimension, since real-life BCI applications on ASD patients pose several challenges, this dataset provide a test bench for data quality and artifactual EEG data on ASD population that new projects can use to validate its models before approaching the clinical patients directly.

CONCLUSION

This paper presented the BCIAUT-P300 dataset which combines multi-session and multi-subject data of 15 ASD participants using a P300-based BCI for training joint-attention skills. The dataset was used on the IFMBE scientific competition where 9 teams from around the world reach the final phase and presented their methods, which were briefly presented here. Overall, deep learning methods were able to overcome the more traditional machine learning approaches, with the best method obtaining an average accuracy of 92.3%. Future studies should address the multiple dimensions of the dataset to reduce training times while improving accuracy.

MEMBERS OF THE GBT-UPM AND NEURAL_ENGINEERING_GROUP

GBT-UPM: Lucía de Arancibia, Patricia Sánchez-González, Enrique J. Gómez, M. Elena Hernando, and Ignacio Oropesa.

Neural_Engineering_Group: Bipra Chatterjee, Ramaswamy Palaniappan, and Cota Navin Gupta.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.kaggle.com/disbeat/bciaut-p300 (doi: 10.34740/kaggle/dsv/1375326).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by CEIC—Comissão de Ética para a Investigação Clínica (Portuguese Ethics Committee for Clinical Research). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

MS organized the challenge, the dataset and partially wrote most of the sections of the document along with DB. DB also performed the statistical analysis. DB, ES-V, GBT-UPM, MB-V, DK, AM, NEG, TS, and HZ wrote the ID-1:9 descriptions, in that order. CA was responsible for the BCI system development and data collection and, along with MS, BD, and JH coordinated the IFMBE Scientific Challenge. PC supervised the IFMBE

Scientific Challenge and MC-B was the PI of the BCIAUT clinical trial, providing scientific guidance to all the process. All authors reviewed and made significant contributions to the final document.

FUNDING

This work was supported by Fundação para a Ciência e Tecnologia, PAC –286 MEDPERSYST, POCI-01-0145-FEDER-016428, BIGDATIMAGE, CENTRO-01-0145-FEDER-000016 financed by Centro 2020 FEDER, COMPETE, FCT UID/4950/2020, PTDC/PSI-GER/30852/2017. For Neural Engineering Group: Bipra Chatterjee was funded by MHRD Doctoral Student Scholarship from the Government of India and Cota Navin Gupta's time was funded by Scheme for Promotion of Academic and Research Collaboration (SPARC Grant), Project Code: P1073.

ACKNOWLEDGMENTS

The Scientific Challenge is an initiative sponsored by the Health Informatics and eHealth Working Group of the International Federation on Medical and Biological Engineering – IFMBE.

REFERENCES

Adama, V. S., Schindler, B., and Schmid, T. (2020). "Using time domain and pearson's correlation to predict attention focus in autistic spectrum disorder from EEG P300 components," in *XV Mediterranean Conference on Medical and Biological Engineering and Computing – MEDICON 2019*, eds J. Henriques, N. Neves, and P. de Carvalho (Cham: Springer), 1890–1893. doi: 10.1007/978-3-030-31635-8_230

Adamson, L. B., Bakeman, R., Suma, K., and Robins, D. L. (2019). An expanded view of joint attention: skill, engagement, and language in typical development and autism. *Child Dev.* 90, e1–e18. doi: 10.1111/cdev.12973

Agapov, S. N., Bulanov, V. A., Zakharov, A. V., and Sergeeva, M. S. (2016). *Wavelet Algorithm for the Identification of P300 ERP Component*. Available online at: http://arxiv.org/abs/1611.00033 (accessed May 10, 2020).

Amaral, C., Mouga, S., Simões, M., Pereira, H. C., Bernardino, I., Quental, H., et al. (2018). A feasibility clinical trial to improve social attention in autistic spectrum disorder (ASD) using a brain computer interface. *Front. Neurosci.* 12:477. doi: 10.3389/fnins.2018.00477

Amaral, C., Simões, M., Souga, S., Andrade, J., and Castelo-Branco, M. (2017). A novel Brain Computer Interface for classification of social joint attention in autism and comparison of 3 experimental setups: a feasibility study. *J. Neurosci. Methods* 290, 105–115. doi: 10.1016/j.jneumeth.2017.07.029

Bamdad, M., Zarshenas, H., and Auais, M. A. (2015). Application of BCI systems in neurorehabilitation: a scoping review. *Disabil. Rehabil. Assist. Technol.* 10, 355–364. doi: 10.3109/17483107.2014.961569

Barachant, A., Bonnet, S., Congedo, M., and Jutten, C. (2010). "Riemannian geometry applied to BCI classification," in *Latent Variable Analysis and Signal Separation. LVA/ICA 2010. Lecture Notes in Computer Science*, eds V. Vigneron, V. Zarzoso, E. Moreau, R. Gribonval, and E. Vincent (Berlin: Springer), 629–636. doi: 10.1007/978-3-642-15995-4_78

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Bhattacharyya, S., Khasnobish, A., Ghosh, P., Mazumder, A., and Tibarewala, D. N. (2017). "A review on brain imaging techniques for BCI applications," in *Medical Imaging: Concepts, Methodologies, Tools, and Applications*, ed. Information Resources Management Association (Pennsylvania: IGI Global), 300–330. doi: 10.4018/978-1-5225-0571-6.ch012

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Cham: Springer.

Bittencourt-Villalpando, M., and Maurits, N. M. (2018). Stimuli and feature extraction algorithms for brain-computer interfaces: a systematic comparison. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 1669–1679. doi: 10.1109/tnsre.2018.2855801

Bittencourt-Villalpando, M., and Maurits, N. M. (2020). "Linear SVM algorithm optimization for an EEG-based brain-computer interface used by high functioning autism spectrum disorder participants," in *XV Mediterranean Conference on Medical and Biological Engineering and Computing – MEDICON 2019*, eds J. Henriques, N. Neves, and P. de Carvalho (Cham: Springer), 1875–1884. doi: 10.1007/978-3-030-31635-8_228

Blankertz, B., Müller, K. R., Curio, G., Vaughan, T. M., Schalk, G., Wolpaw, J. R., et al. (2004). The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials. *IEEE Trans. Biomed. Eng.* 51, 1044–1051. doi: 10.1109/TBME.2004.826692

Blankertz, B., Müller, K. R., Krusienski, D. J., Schalk, G., Wolpaw, J. R., Schlögl, A., et al. (2006). The BCI competition III: validating alternative approaches to actual BCI problems. *IEEE Trans. Neural Syst. Rehabil. Eng.* 14, 153–159. doi: 10.1109/TNSRE.2006.875642

Borra, D., Fantozzi, S., and Magosso, E. (2020a). "Convolutional neural network for a P300 brain-computer interface to improve social attention in autistic spectrum disorder," in *XV Mediterranean Conference on Medical and Biological Engineering and Computing – MEDICON 2019*, eds J. Henriques, N. Neves, and P. de Carvalho (Cham: Springer), 1837–1843. doi: 10.1007/978-3-030-31635-8_223

Borra, D., Fantozzi, S., and Magosso, E. (2020b). Interpretable and lightweight convolutional neural network for EEG decoding: application to movement execution and imagination. *Neural Netw.* 129, 55–74. doi: 10.1016/j.neunet.2020.05.032

Bostanov, V., and Kotchoubey, B. (2006). The t-CWT: a new ERP detection and quantification method based on the continuous wavelet transform and Student's t-statistics. *Clin. Neurophysiol.* 117, 2627–2644. doi: 10.1016/j.clinph.2006.08.012

Chatterjee, B., Palaniappan, R., and Gupta, C. N. (2020). "Performance evaluation of manifold algorithms on a P300 paradigm based online BCI dataset," in *XV Mediterranean Conference on Medical and Biological Engineering and Computing – MEDICON 2019*, eds J. Henriques, N. Neves, and P. de Carvalho (Cham: Springer), 1894–1898. doi: 10.1007/978-3-030-31635-8_231

Chaudhary, U., Birbaumer, N., and Ramos-Murguialday, A. (2016). Brain-computer interfaces for communication and rehabilitation. *Nat. Rev. Neurol.* 12, 513–525. doi: 10.1038/nrneurol.2016.113

Chollet, F. (2016). "Xception: deep learning with depthwise separable convolutions," in *Proceedings of the 2017 IEEE Conf. Comput. Vis. Pattern Recognit.*,(Piscataway, NJ: IEEE), 1800–1807.

Craik, A., He, Y., and Contreras-Vidal, J. L. (2019). Deep learning for electroencephalogram (EEG) classification tasks: a review. *J. Neural Eng.* 16:031001. doi: 10.1088/1741-2552/ab0ab5

de Arancibia, L., Sánchez-González, P., Gómez, E. J., Hernando, M. E., and Oropesa, I. (2020). "Linear vs nonlinear classification of social joint attention in autism using VR P300-based brain computer interfaces," in *XV Mediterranean Conference on Medical and Biological Engineering and Computing – MEDICON 2019*, eds J. Henriques, N. Neves, and P. de Carvalho (Cham: Springer), 1869–1874. doi: 10.1007/978-3-030-31635-8_227

Demiralp, T., Ademoglu, A., Istefanopulos, Y., Başar-Eroglu, C., and Başar, E. (2001). Wavelet analysis of oddball P300. *Int. J. Psychophysiol.* 39, 221–227. doi: 10.1016/S0167-8760(00)00143-4

Deshpande, G., Rangaprakash, D., Oeding, L., Cichocki, A., and Hu, X. P. (2017). A new generation of brain-computer interfaces driven by discovery of latent EEG-fMRI linkages using tensor decomposition. *Front. Neurosci.* 11:246. doi: 10.3389/fnins.2017.00246

Drugowitsch, J. (2013). Variational Bayesian inference for linear and logistic regression. arXiv [Preprint]. Avaialble at: https://arxiv.org/abs/1310.5438 (accessed May 10, 2020).

Farwell, L., and Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electr. Clin. Neurophysiol.* 70, 510–523. doi: 10.1016/0013-4694(88)90149-6

Guo, M., Jin, J., Jiao, Y., Wang, X., and Cichockia, A. (2019). Investigation of visual stimulus with various colors and the layout for the oddball paradigm in evoked related potential-based brain–computer interface. *Front. Comput. Neurosci.* 13:24. doi: 10.3389/fncom.2019.00024

Korczowski, L., Congedo, M., and Jutten, C. (2015). "Single-trial classification of multi-user P300-based brain-computer interface using riemannian geometry," in *Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (Piscataway, NJ: IEEE), 1769–1772.

Krusienski, D. J., Sellers, E. W., Cabestaing, F., Bayoudh, S., McFarland, D. J., Vaughan, T. M., et al. (2006). A comparison of classification techniques for the P300 Speller. *J. Neural Eng.* 3:299.

Krzemiński, D., Michelmann, S., Treder, M., and Santamaria, L. (2020). "Classification of P300 component using a riemannian ensemble approach," in *XV Mediterranean Conference on Medical and Biological Engineering and Computing – MEDICON 2019*, eds J. Henriques, N. Neves, and P. de Carvalho (Cham: Springer), 1885–1889. doi: 10.1007/978-3-030-31635-8_229

Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., and Lance, B. J. (2018). EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. *J. Neural Eng.* 15:aace8c. doi: 10.1088/1741-2552/aace8c

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521:436.

Lopes, A. C., Rodrigues, J., Perdigão, J., Pires, G., and Nunes, U. J. (2016). A new hybrid motion planner. *IEEE Robot. Autom. Mag.* 23, 82–93. doi: 10.1109/MRA.2016.2605403

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., et al. (2018). A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update. *J. Neural Eng.* 15, R1–R13. doi: 10.1088/1741-2552/aab2f2

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., and Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain-computer interfaces. *J. Neural Eng.* 4, R1–R13.

MacKay, D. J. C. (1992). Bayesian interpolation. *Neural Comput.* 4, 415–447. doi: 10.1162/neco.1992.4.3.415

Manor, R., and Geva, A. B. (2015). Convolutional neural network for multi-category rapid serial visual presentation BCI. *Front. Comput. Neurosci.* 9:146.

McFarland, D. J., and Wolpaw, J. R. (2017). EEG-based brain–computer interfaces. *Curr. Opin. Biomed. Eng.* 4, 194–200. doi: 10.1016/j.cobme.2017.11.004

Miladinović, A., Ajčević, M., Battaglini, P. P., Silveri, G., Ciacchi, G., Morra, G., et al. (2020). "Slow cortical potential BCI classification using sparse variational bayesian logistic regression with automatic relevance determination," in *Proceedings of the XV Mediterranean Conference on Medical and Biological Engineering and Computing – MEDICON 2019*, eds J. Henriques, N. Neves, and P. de Carvalho (Cham: Springer), 1853–1860. doi: 10.1007/978-3-030-31635-8_225

Nakanishi, M., Wang, Y.-T., Wei, C.-S., Chiang, K.-J., and Jung, T.-P. (2019). Facilitating calibration in high-speed BCI spellers via leveraging cross-device shared latent responses. *IEEE Trans. Biomed. Eng.* 67, 1105–1113. doi: 10.1109/tbme.2019.2929745

Pinegger, A., Hiebel, H., Wriessnegger, S. C., and Müller-Putz, G. R. (2017). Composing only by thought: novel application of the P300 brain-computer interface. *PLoS One* 12:e0181584. doi: 10.1371/journal.pone.0181584

Rakotomamonjy, A., and Guigue, V. (2008). BCI competition III: dataset II-ensemble of SVMs for BCI P300 speller. *IEEE Trans. Biomed. Eng.* 55, 1147–1154. doi: 10.1109/TBME.2008.915728

Raza, H., Rathee, D., Zhou, S. M., Cecotti, H., and Prasad, G. (2019). Covariate shift estimation based adaptive ensemble learning for handling non-stationarity in motor imagery related EEG-based brain-computer interface. *Neurocomputing* 343, 154–166. doi: 10.1016/j.neucom.2018.04.087

Riggins, T., and Scott, L. S. (2019). P300 development from infancy to adolescence. *Psychophysiology* 57:e13346. doi: 10.1111/psyp.13346

Saha, S., and Baumert, M. (2020). Intra- and inter-subject variability in EEG-based sensorimotor brain computer interface: a review. *Front. Comput. Neurosci.* 13:87. doi: 10.3389/fncom.2019.00087

Sajda, P., Gerson, A., Muller, K.-R., Blankertz, B., and Parra, L. (2003). A data analysis competition to evaluate machine learning algorithms for use in brain-computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* 11, 184–185. doi: 10.1109/TNSRE.2003.814453

Santamaría-Vázquez, E., Martínez-Cagigal, V., Gomez-Pilar, J., and Hornero, R. (2020). "Deep learning architecture based on the combination of convolutional and recurrent layers for ERP-based brain-computer interfaces," in *XV Mediterranean Conference on Medical and Biological Engineering and Computing – MEDICON 2019*, eds J. Henriques, N. Neves, and P. de Carvalho (Cham: Springer), 1844–1852. doi: 10.1007/978-3-030-31635-8_224

Simões, M., Amaral, C., França, F., Carvalho, P., and Castelo-Branco, M. (2019). "Applying weightless neural networks to a P300-based brain-computer interface," in *World Congress on Medical Physics and Biomedical Engineering 2018. IFMBE Proceedings*, eds L. Lhotska, L. Sukupova, I. Lackoviæ, and G. Ibbott (Singapore: Springer), 113–117. doi: 10.1007/978-981-10-9023-3_20

Sitaram, R., Caria, A., Veit, R., Gaber, T., Rota, G., Kuebler, A., et al. (2007). FMRI brain-computer interface: a tool for neuroscientific research and treatment. *Comput. Intell. Neurosci.* 2007:25487. doi: 10.1155/2007/25487

Tangermann, M., Müller, K. R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., et al. (2012). Review of the BCI competition IV. *Front. Neurosci.* 6:55. doi: 10.3389/fnins.2012.00055

Wolpaw, J. R., and Winter Wolpaw, E. (2012). "Brain–computer interfaces: something new under the sun," in *Brain–Computer InterfacesPrinciples and Practice*, eds J. R. Wolpaw and E. W. Wolpaw (Oxford: Oxford University Press), 3–12. doi: 10.1093/acprof:oso/9780195388855.003.0001

Yger, F., Berar, M., and Lotte, F. (2017). Riemannian approaches in brain-computer interfaces: a review. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 1753–1762. doi: 10.1109/TNSRE.2016.2627016

Zhang, X., Yao, L., Wang, X., Monaghan, J., Mcalpine, D., and Zhang, Y. (2019). A survey on deep learning based brain computer interface: recent advances and new frontiers. *ACM Int. Conf. Proc. Ser.* 1, 504–513.

Zhao, D., Tang, F., Si, B., and Feng, X. (2019). Learning joint space–time–frequency features for EEG decoding on small labeled data. *Neural Netw.* 114, 67–77. doi: 10.1016/j.neunet.2019.02.009

Zhao, H., Yu, S., Prinable, J., McEwan, A., and Karlsson, P. (2020). "A feasible classification algorithm for event-related potential (ERP) based brain-computer-interface (BCI) from IFMBE scientific challenge dataset," in *XV Mediterranean Conference on Medical and Biological Engineering and Computing – MEDICON 2019*, eds J. Henriques, N. Neves, and P. de Carvalho (Cham: Springer), 1861–1868. doi: 10.1007/978-3-030-31635-8_226

Zou, Y., Zhao, X., Chu, Y., Zhao, Y., Xu, W., and Han, J. (2019). An inter-subject model to reduce the calibration time for motion imagination-based brain-computer interface. *Med. Biol. Eng. Comput.* 57, 939–952. doi: 10.1007/s11517-018-1917-x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

DATA REPORT
published: 29 September 2020
doi: 10.3389/fnins.2020.579353

Open Access Multimodal fNIRS Resting State Dataset With and Without Synthetic Hemodynamic Responses

Alexander von Lühmann [1], Xinge Li [1], Natalie Gilmore [2], David A. Boas [1] and Meryem A. Yücel [1]**

[1] Neurophotonics Center, Biomedical Engineering, Boston University, Boston, MA, United States, [2] Department of Speech, Language and Hearing Sciences, Boston University, Boston, MA, United States

Keywords: multimodal, open access, synthetic HRF, resting, fNIRS

OPEN ACCESS

Edited by:
Ian Daly,
University of Essex, United Kingdom

Reviewed by:
Vassiliy Tsytsarev,
University of Maryland, College Park, United States
Wei-Peng Teo,
Nanyang Technological University, Singapore

****Correspondence:***
Alexander von Lühmann
avolu@bu.edu
Meryem A. Yücel
mayucel@bu.edu

Specialty section:
This article was submitted to Neural Technology, a section of the journal Frontiers in Neuroscience

Received: *02 July 2020*
Accepted: *19 August 2020*
Published: *29 September 2020*

Citation:
von Lühmann A, Li X, Gilmore N, Boas DA and Yücel MA (2020) Open Access Multimodal fNIRS Resting State Dataset With and Without Synthetic Hemodynamic Responses. Front. Neurosci. 14:579353. doi: 10.3389/fnins.2020.579353

INTRODUCTION

Functional Near-Infrared Spectroscopy (fNIRS) is an optical neuroimaging technology that has rapidly gained momentum within the last decades (Boas et al., 2014; Scholkmann et al., 2014; Yücel et al., 2017). It is a non-hazardous and non-invasive optical brain imaging technique that uses near-infrared light to measure local cortical concentration changes of oxygenated and deoxygenated hemoglobin (HbO_2/HbR), which are associated with brain metabolism (Villringer and Chance, 1997; Ferrari and Quaresima, 2012). fNIRS has been considered a cost-effective and mobile alternative for functional Magnetic Resonance Imaging in conventional neuroscientific research. It is very suitable—and thus increasingly being used—for single trial analysis and Brain Computer Interface (BCI) applications (Matthews et al., 2008; Hong et al., 2018) as a single modality or along with Electroencephalography (EEG). While EEG and fNIRS signal processing is essential to increase the contrast to noise ratio (CNR) of measured brain responses, the nature of the signals and processing methods differ greatly. Hemodynamic brain responses in fNIRS are usually masked by local and systemic physiological confounding signals, for instance from superficial (scalp) blood flow, low frequency oscillations (Mayer waves), motion and breathing (Elwell et al., 1999; Yücel et al., 2016; von Lühmann et al., 2019). New and increasingly complex and powerful statistical methods are being developed that aim to remove the confounding factors in the signal, improve CNR and increase the detection/classification accuracy of hemodynamic responses. An objective way of validating the power of these novel methods and comparing them with the existing ones is to use an fNIRS dataset which has all the confounding signals but also a known hemodynamic brain response. One solution for this problem is to generate realistic fNIRS ground truth data by modeling a hemodynamic response function (HRF) on top of real resting state data (Gagnon et al., 2012; von Lühmann et al., 2019, 2020a,b). This approach can be used as a good approximation for realistic fNIRS signals with evoked responses, for which the ground truth is available. Generating such data is comparatively straight forward but requires prior knowledge in fNIRS signal characteristics as well as experience in fNIRS signal processing. Moreover, the use of short-separation fNIRS measurements and additional physiological signals, such as accelerometer or photoplethysmography (PPG), has been shown to enable methods that yield improved CNR (Yücel et al., 2015; von Lühmann et al., 2020a), but there are only few openly available multimodal fNIRS datasets (Shin et al., 2017, 2018) and even fewer multimodal datasets that include sufficient resting state periods to enable the approach described above. Thus, as a remedy, here we provide such a multimodal dataset with (and without) added synthetic HRF ground truth,

short-separation fNIRS measurements, accelerometer, and other physiological measurement, for the data science community in order to facilitate the validation of novel methods. We also provide a simple code example to enable customization and modification of the HRF ground truth in the data.

METHODS

The resting state data consists of two subsets: Dataset I, with 5 min resting state data from 14 participants and Dataset II, with 10 min resting state data from 14 participants. The data details follow and are summarized in **Figure 1**.

Participants and Demographics

Dataset I consists of recordings from 14 healthy participants (age: 21 ± 2 years; 11 male/3 female) and Dataset II from 14 healthy participants (age: 32 ± 19 years; 7 male/6 female/1 not reported) with no neurological or psychological disorders.

Experimental Paradigm

Participants were seated in a comfortable chair with room light on and were asked to look at a fixation cross on a black screen ∼50 cm in front of them. A 5-min resting state data (Dataset I) or a 10-min resting state data (Dataset II) were recorded from each participant.

Data Acquisition

fNIRS data were acquired using a multichannel continuous wave fNIRS system (CW6, TechEn Inc. MA, USA) operating at 690 and 830 nm wavelengths. The system is an optical imager with 32 frequency encoded lasers (half at 690 and half at 830 nm) and 32 avalanche photo-diode detectors. The light is carried from the system to the head probe and back via optical fiber bundles. fNIRS data were acquired at a sample rate of 50 Hz.

DATASET I: Optode Array and Auxiliary Measurements

Optode array

Both head optode arrays were designed utilizing AtlasViewer software (Aasted et al., 2015) (**Figure 1**). The optode array for Dataset I consisted of an elastic cap (EasyCap, Herrsching, Germany) with 8 sources, 12 long-separation detectors (∼3 cm apart from the source) and 2 short-separation detectors (∼1 cm apart from the source) providing, in total, 26 long-separation and 2 short-separation channels covering the occipital lobe.

Auxiliary measurements

Systemic physiological changes and head motions of the participants were simultaneously recorded along with the fNIRS data using an MP160 data acquisition and analysis system (BIOPAC Systems Inc., Goleta, CA). The pulse waveform was recorded using a PPG100C amplifier and TSD200 PPG pulse transducer placed on the participant's right index finger (BIOPAC Systems Inc., Goleta, CA). Respiration data was collected via measuring the abdominal (or thoracic) expansion and contraction using a RSP100C amplifier and a TSD201 respiration transducer (respiration belt) (BIOPAC Systems Inc., Goleta, CA) around the participant's chest. The blood pressure waveform was recorded using a DA100C amplifier and a TSD110 pressure transducer (BIOPAC Systems Inc., Goleta, CA) placed on the participant's right thumb. Head motions in x, y, z directions were collected using an accelerometer (ADXL335, Analog Devices Inc., Norwood, MA) secured on the head with a headband.

DATASET II: Optode Array and Auxiliary Measurements

Optode array

The optode array for Dataset II consisted of an elastic cap (EasyCap, Herrsching, Germany) with 16 sources, 24 long-separation detectors (∼3 cm apart from the source) and 8 short-separation detectors (∼1 cm apart from the source) providing, in total, 48 long-separation and 8 short-separation channels covering the head from frontal to parietal regions bilaterally.

Auxiliary measurement

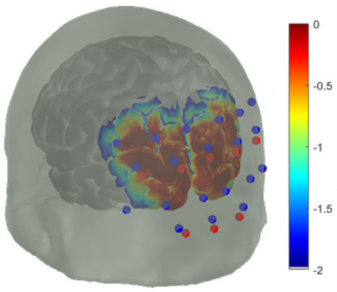
Head motions of the participants in x, y, z directions were simultaneously recorded along with the fNIRS data using a 3-axis accelerometer (ADXL335, Analog Devices Inc., Norwood, MA) secured on the head with a headband.

Adding Synthetic HRF to the fNIRS Data

In the documented data repository, we provide the acquired resting state data with and without synthetic HRF as well as the scripts used for the generation of the data to enable users to alter and re-generate ground truth HRF according to their needs. We generate synthetic HRFs with three different amplitudes using a gamma function with a time-to-peak of 6 s and a total duration of 16.5 s. The shape of this synthetic HRF is also depicted in **Figure 2A**. The three amplitudes are provided as percentages (100/50/20%) of a typical average amplitude of a task-evoked HRF (Huppert et al., 2006) and simulate varying degrees of CNR in the data: The (100%) amplitude is equal to +1% | −2% change from baseline intensity at 690 nm | 830 nm leading to an HRF peak amplitude of +0.66 | −0.23 μM for HbO_2 | HbR, respectively with a differential pathlength factor of 6 (Delpy et al., 1988; Boas et al., 2004) for a 30 mm source-detector separation. For each participant in the two datasets, all resting state data is divided into windows of 20 s length. The HRFs are added in the intensity domain at a random onset (0–3.5 s) within each 20 s window for a randomly selected half of all available long separation channels after pruning with a 5 dB SNR threshold. This results in an average of 15 | 38 trials per participant and HRF amplitude in each resting state Dataset I | II.

Data Structure and Format

Both datasets are presented in the Shared Near Infrared File Format V1.0 Specification (snirf), which is based on the HDF5 file format (https://github.com/fNIRS/snirf). ***SnirfClass*** function loads the snirf object into the MATLAB environment. **Table 1** provides the list of variables in the current dataset snirf object. The main fields of interest are: the ***data*** field which has the fNIRS raw signal at each channel



Set I	Resting State Data	Set II
14	# of participants	14
21 ± 2	age [years]	30 ± 17
26 + 2	# channels (LS+SS)	48 + 8
5	duration [min]	10
occipital	optode array (coverage)	fronto-parietal
PPG, RESP, BP, ACCEL	auxiliary measurements	ACCEL
20, 50, 100%	HRF amplitudes	20, 50, 100%
15	# trials / recording	38

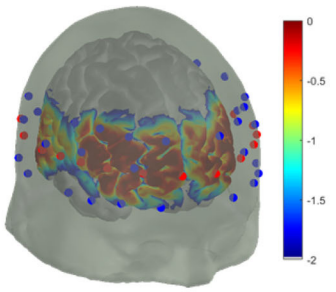


FIGURE 1 | Summary of dataset metadata. Red dots are fNIRS emitters, blue dots are fNIRS detectors. LS, Long-Separation; SS, Short-Separation; PPG, Photoplethysmography; RESP, Respiration; BP, Blood Pressure; ACCEL, Accelerometer. HRF with 20, 50, and 100 % of +0.66 | −0.23 μM peak HbO_2/HbR amplitude. Sensitivity profile in log10 dB.

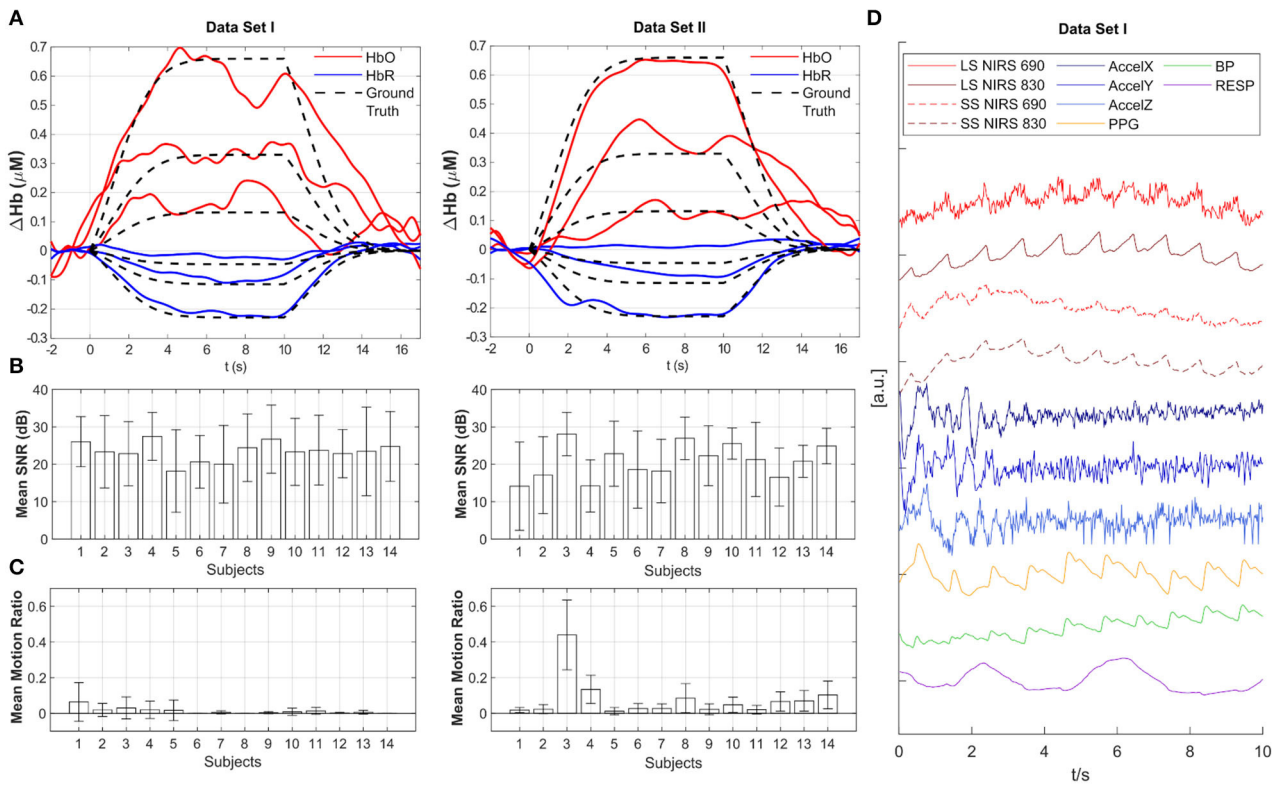


FIGURE 2 | Overview of data quality and baseline analysis for Datasets I and II. **(A)** Across-trial recovered HRF estimated with tCCA GLM using Homer3 for all three amplitudes (100, 50, 20%) (participant 33 from Dataset I and participant 98 from Dataset II - the first participant in each data set). Red: HbO_2, Blue: HbR, Black dashed: Ground Truth. **(B)** Mean SNR in dB across channels for all participants. Whiskers indicate standard deviation. **(C)** Mean Motion Ratio across channels for all participants which indicates the ratio of motion-contaminated data to the whole data. **(D)** Example time course of multimodal data in Dataset I. LS, Long-separation channel; SS, Short-separation channel; AccelX|Y|Z, Accelerometer; PPG, Photoplethysmogram; BP, Blood Pressure; RESP, Respiration.

and relevant information, the ***probe*** field which has optode array information, the ***aux*** field which has all the auxiliary measurements and their details, and the ***stim*** field which has the experimental paradigm information. Please note that, while snirf.aux(2)/(3)/(4) have AccelX, AccelY, and AccelZ measurements for both datasets, Dataset I has PPG, blood

TABLE 1 | Snirf object fields.

snirf.filename	Filename
snirf.fileformat	"hdf5"
snirf.data.dataTimeSeries	Time-varying signals from all channels following the order in **snirf.data.measurementList** e.g., **10th** column of **snirf.data.dataTimeSeries** corresponds to **snirf.data.measurementList(10)** which is the channel defined with sourceIndex: 2; detectorIndex: 18 and wavelengthIndex: 1
snirf.data.time	Time
snirf.data.measurementList	Per-channel source-detector-wavelength information
snirf.data.measurementList.dataTypeLabel	Defined as (1) for HRF added channels and (0) for no HRF channels, specifically for this dataset e.g., to check whether HRF is added on channel 10 check **snirf.data.measurementList(10).dataTypeLabel**
snirf.stim.name	Stimuli labels
snirf.stim.data	Data stream of the stimulus channel
snirf.probe.wavelengths	List of wavelengths (in nm)
snirf.probe.sourcePos	Source position
snirf.probe.detectorPos	Detector position
snirf.probe.sourceLabels	String arrays specifying source names
snirf.probe.detectorLabels	String arrays specifying detector names
snirf.aux.name	Name of the auxiliary channel e.g., to check the content of an auxiliary channel 2 **snirf.aux(2).name**
snirf.aux.dataTimeSeries	Data acquired from the auxiliary channel
snirf.aux.time	Time for auxiliary data

pressure (BP), RESP at snirf.aux(5), (6), and (7), respectively in addition to these.

BASELINE ANALYSIS AND DATA QUALITY ASSESSMENT

Baseline analysis and data quality assessment was performed using the openly available Homer3 toolbox (https://github.com/BUNPC/Homer3) (Huppert et al., 2009). HRFs were recovered from the augmented resting state data using the processing stream provided in the repository (***tCCA_xmpl_procStream_Gauss_noHPF.cfg*** *under* ***"code"*** *folder*). This processing stream includes 0.5 Hz zero phase low pass filter with an effective order of 6, conversion to HbO_2 and HbR using the modified Beer-Lambert Law (Delpy et al., 1988; Boas et al., 2004), and subsequent HRF estimation with the temporally embedded General Linear Model (tCCA GLM) approach using short-distance channels and a polynomial drift term for nuisance regression and Gaussian basis functions for the HRF regressor (von Lühmann et al., 2020a). **Figure 2A** exemplifies the resulting HRF estimates in one augmented channel for all three amplitudes from participant 33 from Dataset I and participant 98 from Dataset II. Data quality is provided for each participant as across-channel average of the Signal to Noise Ratio (SNR) in **Figure 2B** and as the mean motion ratio across channels in **Figure 2C.** Channel SNR is calculated as 20 × log10 of the mean over std. of the raw intensity signal. The motion ratio is calculated as the ratio between the cumulative time of segments in the data that were considered to be confounded by motion artifacts, as identified by the Homer2 function ***hmrMotionArtifactByChannel*** (with tMotion = 0.5, tMask = 0.5, STDEVthresh = 20, AMPthresh = 5), to the total acquisition time. **Figure 2D** displays a typical segment of all available signals (z-scored) in the first participant in Dataset I. Long and short-separation fNIRS channels exhibit typical low frequency components and cardiac pulsation, which is also present in the PPG and BP measurement.

SUMMARY

We reported a multimodal fNIRS resting state dataset from 28 participants, that we provide with and without added synthetic HRF ground truth at three different amplitudes. We include the script used for the generation of these data to enable users to adapt this approach to their own needs. The availability of multiple auxiliary biosignals, such as motion (accelerometer) and PPG in the data, can be used to explore and extend existing multimodal fNIRS-based signal processing approaches (von Lühmann et al., 2019, 2020a). Resting fNIRS data with added known HRF enables the validation of novel processing methods for single trial HRF detection and BCI as well as more general artifact rejection and preprocessing approaches and their comparison with existing methods. This can also be useful for methods that tackle challenges such as non-stationarities in the amplitude and time to peak of hemodynamic responses to a stimulus.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data and the code for HRF ground-truth generation can be found here: https://www.nitrc.org/frs/?group_id=1071.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Boston University IRB. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

XL and NG collected the first and second datasets, respectively. AL and MY analyzed the data, prepared the figures, and wrote the report. DB reviewed the report. All authors confirmed the final version of the report.

FUNDING

This work was funded by a research contract under Facebook's Sponsored Academic Research Agreement and in part by NIH R24NS104096.

REFERENCES

Aasted, C. M., Yücel, M., A., Cooper, R. J., Dubb, J., Tsuzuki, D., Becerra, L., et al. (2015). Anatomical guidance for functional near-infrared spectroscopy: AtlasViewer tutorial. *Neurophotonics* 2:020801. doi: 10.1117/1.NPh.2.2.020801

Boas, D. A., Dale, A. M., and Franceschini, M. A. (2004). Diffuse optical imaging of brain activation: approaches to optimizing image sensitivity, resolution, and accuracy. *Neuroimage* 23, 275–288. doi: 10.1016/j.neuroimage.2004.07.011

Boas, D. A., Elwell, C. E., Ferrari, M., and Taga, G. (2014). Twenty years of functional near-infrared spectroscopy: introduction for the special issue. *Neuroimage* 85, 1–5. doi: 10.1016/j.neuroimage.2013.11.033

Delpy, D. T., Cope, M., van der Zee, P., Arridge, S., Wray, S., and Wyatt, J. (1988). Estimation of optical pathlength through tissue from direct time of flight measurement. *Phys. Med. Biol.* 33:1433. doi: 10.1088/0031-9155/33/12/008

Elwell, C. E., Springett, R., Hillman, E., and Delpy, D. T. (1999). Oscillations in cerebral haemodynamics. Implications for functional activation studies. *Advances in. Experimental Exp. Medicine and. Biology Biol. vol.* 471, 57–65. doi: 10.1007/978-1-4615-4717-4_8

Ferrari, M., and Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *Neuroimage* 63, 921–935. doi: 10.1016/j.neuroimage.2012.03.049

Gagnon, L., Cooper, R. J., Yücel, M. A., Perdue, K. L., Greve, D. N., and Boas, D. A. (2012). Short separation channel location impacts the performance of short channel regression in NIRS. *Neuroimage* 59, 2518–2528. doi: 10.1016/j.neuroimage.2011.08.095

Hong, K.-S., Khan, M. J., and Hong, M. J. (2018). Feature extraction and classification methods for hybrid fNIRS-EEG brain-computer interfaces. *Front. Hum. Neurosci.* 12:246. doi: 10.3389/fnhum.2018.00246

Huppert, T. J., Diamond, S. G., Franceschini, M. A., and Boas, D. A. (2009). HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Appl. Opt.* 48, 280–298. doi: 10.1364/AO.48.00D280

Huppert, T. J., Hoge, R. D., Diamond, S. G., Franceschini, M. A., and Boas, D. A. (2006). A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *Neuroimage* 29, 368–382. doi: 10.1016/j.neuroimage.2005.08.065

Matthews, F., Pearlmutter, B. A., Ward, T. E., Soraghan, C., and Markham, C. (2008). Hemodynamics for brain-computer interfaces. *Signal Process. Mag. IEEE* 25, 87–94. doi: 10.1109/MSP.2008.4408445

Scholkmann, F., Kleiser, S., Metz, A. J., Zimmermann, R., Mata Pavia, J., Wolf, U., et al. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *Neuroimage* 85, 6–27. doi: 10.1016/j.neuroimage.2013.05.004

Shin, J., von Lühmann, A., Blankertz, B., Kim, D.-W. W., Jeong, J., Hwang, H.-J. J., et al. (2017). Open access dataset for EEG+ NIRS single-trial classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 1735–1745. doi: 10.1109/TNSRE.2016.2628057

Shin, J., Von Lühmann, A., Kim, D. W., Mehnert, J., Hwang, H. J., and Müller, K. R. (2018). Data descriptor: simultaneous acquisition of EEG and NIRS during cognitive tasks for an open access dataset. *Sci. Data* 5:180003. doi: 10.1038/sdata.2018.3

Villringer, A., and Chance, B. (1997). Non-invasive optical spectroscopy and imaging of human brain function. *Trends Neurosci.* 20, 435–442. doi: 10.1016/S0166-2236(97)01132-6

von Lühmann, A., Boukouvalas, Z., Müller, K. R., and Adal,i, T. (2019). A new blind source separation framework for signal analysis and artifact rejection in functional near-infrared spectroscopy. *Neuroimage* 200, 72–88. doi: 10.1016/j.neuroimage.2019.06.021

von Lühmann, A., Li, X., Müller, K. R., Boas, D. A., and Yücel, M. A. (2020a). Improved physiological noise regression in fNIRS: a multimodal extension of the General Linear Model using temporally embedded Canonical Correlation Analysis. *Neuroimage* 208:116472. doi: 10.1016/j.neuroimage.2019.116472

von Lühmann, A., Ortega-Martinez, A., Boas, D. A., and Yücel, M. A. (2020b). Using the general linear model to improve performance in fnirs single trial analysis and classification: a perspective. *Front. Hum. Neurosci.* 14, 1–1730. doi: 10.3389/fnhum.2020.00030

Yücel, M. A., Selb, J., Aasted, C. M., Lin, P.-Y., Borsook, D., Becerra, L., et al. (2016). Mayer waves reduce the accuracy of estimated hemodynamic response functions in functional near-infrared spectroscopy. *Biomed. Opt. Express* 7, 3078–3088. doi: 10.1364/BOE.7.003078

Yücel, M. A., Selb, J., Aasted, C. M., Petkov, M. P., Becerra, L., Borsook, D., et al. (2015). Short separation regression improves statistical significance and better localizes the hemodynamic response obtained by near-infrared spectroscopy for tasks with differing autonomic responses. *Neurophotonics* 2:035005. doi: 10.1117/1.NPh.2.3.035005

Yücel, M. A., Selb, J. J., Huppert, T. J., Franceschini, M. A., and Boas, D. A. (2017). Functional near infrared spectroscopy: enabling routine functional brain imaging. *Curr. Opin. Biomed. Eng.* 4, 78–86. doi: 10.1016/j.cobme.2017.09.011

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The authors declare that this study received funding from Facebook. The funder was involved in the decision to submit the paper for publication.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 02 October 2020
doi: 10.3389/fnins.2020.568000

A Benchmark Dataset for RSVP-Based Brain–Computer Interfaces

Shangen Zhang[1,2], Yijun Wang[3], Lijian Zhang[4] and Xiaorong Gao[5]**

[1] School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing, China, [2] Beijing Key Laboratory of Knowledge Engineering for Materials Science, Beijing, China, [3] State Key Laboratory on Integrated Optoelectronics, Institute of Semiconductors, Chinese Academy of Sciences, Beijing, China, [4] Beijing Machine and Equipment Institute, Beijing, China, [5] Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing, China

This paper reports on a benchmark dataset acquired with a brain–computer interface (BCI) system based on the rapid serial visual presentation (RSVP) paradigm. The dataset consists of 64-channel electroencephalogram (EEG) data from 64 healthy subjects (sub1,. . ., sub64) while they performed a target image detection task. For each subject, the data contained two groups ("A" and "B"). Each group contained two blocks, and each block included 40 trials that corresponded to 40 stimulus sequences. Each sequence contained 100 images presented at 10 Hz (10 images per second). The stimulus images were street-view images of two categories: target images with human and non-target images without human. Target images were presented randomly in the stimulus sequence with a probability of 1∼4%. During the stimulus presentation, subjects were asked to search for the target images and ignore the non-target images in a subjective manner. To keep all original information, the dataset was the raw continuous data without any processing. On one hand, the dataset can be used as a benchmark dataset to compare the algorithms for target identification in RSVP-based BCIs. On the other hand, the dataset can be used to design new system diagrams and evaluate their BCI performance without collecting any new data through offline simulation. Furthermore, the dataset also provides high-quality data for characterizing and modeling event-related potentials (ERPs) and steady-state visual evoked potentials (SSVEPs) in RSVP-based BCIs. The dataset is freely available from http://bci.med.tsinghua.edu.cn/download.html.

Keywords: rapid serial visual presentation, brain–computer interface, electroencephalogram, target detection, public dataset, event-related potential

OPEN ACCESS

Edited by:
Ian Daly,
University of Essex, United Kingdom

Reviewed by:
Yu Zhang,
Stanford University, United States
Jing Jin,
East China University of Science and Technology, China

****Correspondence:***
Yijun Wang
wangyj@semi.ac.cn
Xiaorong Gao
gxr-dea@tsinghua.edu.cn

Specialty section:
This article was submitted to Neural Technology, a section of the journal Frontiers in Neuroscience

Received: *31 May 2020*
Accepted: *08 September 2020*
Published: *02 October 2020*

Citation:
Zhang S, Wang Y, Zhang L and Gao X (2020) A Benchmark Dataset for RSVP-Based Brain–Computer Interfaces. Front. Neurosci. 14:568000. doi: 10.3389/fnins.2020.568000

INTRODUCTION

Brain–computer interfaces (BCIs) provide a direct communication and control channel between the brain and external devices by analyzing neural activity, which has become one of the current study hot spots (Gao et al., 2014; Chen et al., 2015a; Han et al., 2020). Electroencephalogram (EEG) is the most widely used tool for BCIs because of its advantages such as non-invasiveness, low cost, and high temporal resolution (Stegman et al., 2020; Zhang et al., 2020). At present,

remarkable progresses have been made in the performance and practicability of BCIs due to the optimization of the experimental paradigm, the improvement of the signal processing algorithm, and the application of the machine learning method (Chen et al., 2015b; Nakanishi et al., 2018; Zhang et al., 2018). Especially in recent years, the emergence of free open datasets has spared the time, money, and labor costs of data collection, thus providing convenience for the majority of BCI researchers and promoting the progress of algorithm development. The datasets covered many BCI paradigms such as steady-state visual evoked potentials (SSVEPs) (Wang et al., 2017; Lee et al., 2019), event-related P300 potentials (Abibullaev and Zollanvari, 2019; Vaineau et al., 2019), and motor imagery (Cho et al., 2017; Kaya et al., 2018). In addition, there are some open multimodal datasets for BCIs obtained synchronously with EEG (Lioi et al., 2019). As the broad audience of these datasets, researchers in a wide range of fields have contributed their intelligence to the BCI technology.

Rapid serial visual presentation (RSVP)-based BCI is a special type of BCI that detects target stimuli (e.g., letters or images) that are presented sequentially in a stream by detecting the brain's response to the target. RSVP is the process of sequentially displaying images in the same spatial position at a high presentation rate with multiple images per second (such as 2–20 Hz) (Lees et al., 2017). In the applications that benefit from this paradigm, computers are unable to analyze and understand images with deep semantic and unstructured features as successfully as humans, and the manual analysis tools are slow, which makes the study of RSVP-BCI more and more popular in recent decades. RSVP-BCI has been used in counterintelligence, police, and health care that require professionals to review objects, scenes, people, and other relevant information contained in a large number of images (Huang et al., 2017; Singh and Jotheeswaran, 2018; Wu et al., 2018).

Different EEG components are associated with target and non-target stimuli (Bigdely-Shamlo et al., 2008; Cohen, 2014), and BCI signal processing algorithms have been used to recognize event-related potential (ERP) responses and link them to target images. The most commonly exploited ERP in RSVP-based BCI applications is the P300, ideally on a single-trial basis (Manor et al., 2016). In order to detect ERPs induced by target images, researchers have developed a variety of algorithms and evaluated them with the data collected independently (Sajda et al., 2003; Alpert et al., 2014; Zhao et al., 2019). Unfortunately, as far as we know, there is still a lack of a benchmark dataset for the RSVP-based BCI paradigm. It is always difficult to compare the performance of different algorithms with a small amount of data. One of the main difficulties in collecting a benchmark dataset is the large number of system parameters in RSVP-based BCIs (e.g., frequency of image presentation, target definition, target sparsity and identifiability, and number of trials and subjects). There is a great need to collect and publish a large benchmark dataset using the RSVP-based BCI paradigm.

This study provides an open dataset for BCI study based on the RSVP paradigm. The characteristics of this dataset are described as follows. (1) A large number of subjects (64 in total) were recorded. (2) A large number of stimulation image circles (16,000 for each subject) were included. (3) Complete data were provided with the original continuous data without any processing, including EEG data, electrode positions, and subjects information. (4) Stimulus events (onsets and offsets) were precisely synchronized to EEG data. (5) The 64-channel whole-brain EEG data were recorded. That means that this dataset contains a total of 64 subjects, 10,240 trials, 1,024,000 image circles, and 102,400 s of 64-channel EEG data. This dataset provides potential opportunities for developing signal processing and machine learning algorithms that rely on large amounts of EEG data. These features also make it possible to study the algorithms for ERP detection and the methods for stimulus coding with the dataset. In addition, through offline simulation, stimulus coding and target recognition methods can be jointly optimized toward the highest performance of an online BCI.

The rest of this paper is organized as follows. The *Methods* section introduces the experimental setup of data recording. The *Data Recording* section introduces the data records and other relevant information. The *Technical Validations* section introduces the basic methods in data analysis and gives three examples to illustrate how to use the dataset to study the methods of target detection in RSVP-based BCIs. The *Discussions and Conclusion* section summarizes and discusses the future work to improve the dataset.

MATERIALS AND METHODS

Subjects

Sixty-four subjects (32 females; aged 19–27 years, mean age 22 years) with normal or corrected-to-normal vision were recruited for this study. Each subject signed a written informed consent before the experiment and received a monetary compensation for his or her participation. This study was approved by the Research Ethics Committee of Tsinghua University.

Experimental Design

This study developed an offline RSVP-BCI system. A 23.6-inch liquid crystal display (LCD) screen was used to present visual stimuli. The resolution of the screen was 1,920 × 1,080 pixels, and the refresh rate was 60 Hz. The visual stimulus images were rendered within a 1,200 × 800-pixel square in the center of the screen. The screen area surrounding the stimuli image was gray colored [red green blue (RGB): (128, 128, 128)].

The stimulus program was developed under MATLAB (MathWorks, Inc.) using the Psychophysics Toolbox Ver. 3 (PTB-3) (Brainard, 1997). The stimulus images, downloaded from the Computer Science and Artificial Intelligence Library of MIT University, were street-view images of two categories: target images showing human and non-target images without human. During the experiment, subjects were asked to search for the target images and ignore the non-target images in a subjective manner. As previous studies have shown similar performance between motor and non-motor response tasks (Gerson et al., 2006), subjects in this study were required to make a manual button press to maintain attention once detecting target images in the RSVP task.

Figure 1 shows the time course of the RSVP paradigm. Each trial started with a blank for 0.5 s with a cross mark on the center of the screen, and subjects were asked to shift their gaze to the cross mark as soon as possible. The frequency of image presentation was set to 10 Hz (10 images per second).

Figure 2 shows the parameter settings of the experiment for each group. Each group covered two blocks, each containing 40 trials. Each trial contained 100 images, including one, two, three, or four target images. Images in each trial were presented in a random order. At the beginning of each image's presentation, a time marker named "event trigger" was sent by the stimulation program to mark the current stimulus image and was recorded on an event channel of the amplifier synchronized with EEG. There was a short key-controlled pause between trials. The duration of each block was about 10 min. There was an average rest time of about 5 min between two blocks to relieve subjects' fatigue.

Data Acquisition

Electroencephalogram data were recorded using the Synamps2 system (Neuroscan, Inc.) at a sampling rate of 1,000 Hz. All 64 electrodes were used to record EEG and were placed according to the international 10–20 system. The reference electrode, with the 10–20 electrode name of "Ref," was located at the vertex. Electrode impedances were kept below 10 kΩ. During the experiment, subjects were seated in a comfortable chair in a dimly lit soundproof room at a distance of approximately 70 cm from the monitor. The EEG data were filtered from 0.15 to 200 Hz by the system. The power-line noise was removed by a notch filter at 50 Hz. It is worth to mention that the impedance of M1 and M2 electrodes (channels of 33 and 43) was higher than 10 kΩ for some subjects. We therefore suggest to select EEG data from the other 62 channels for analysis, and the EEG analysis in this study used the 62-channel data with the channel indices of [1:32 34:42 44:64] and removed the bad channels.

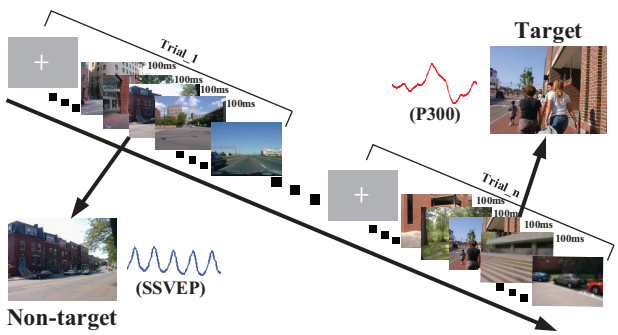


FIGURE 1 | The time course of rapid serial visual presentation (RSVP) paradigm.

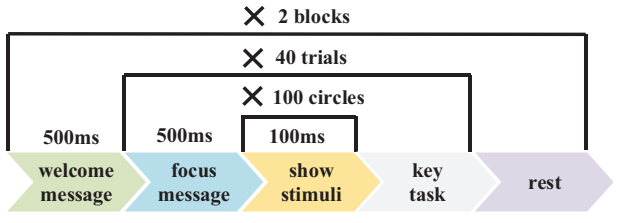


FIGURE 2 | The parameter settings of the experiment for each group.

Data Preprocessing

The dataset was the continuous data at a sampling rate of 250 Hz, and it was obtained from the raw EEG data (sampling rate at 1,000 Hz) after four times downsampling. For each of the datasets from 1 to 64 (sub1,. . ., sub64), EEG data contained four blocks, which were divided into two groups (namely, groups A and B) in chronological order. Each group contained two blocks, and each contained 40 trials. Each trial contained 100 circles, and each circle corresponded to one image. For each group, the two blocks were used for training and testing in the ERP-based target detection, respectively. In addition, a 10-fold cross-validation using both blocks 1 and 2 was performed to further evaluate the classification performance.

To verify the validity of the dataset, the continuous EEG data at a sample rate of 250 Hz were processed by a four-order Butterworth filter with a bandwidth of [2 30] Hz. EEG data epochs were extracted according to event triggers generated by the stimulus program. In this study, time 0 represented the beginning of each image stimulus period (marked by a trigger), and the EEG data corresponding to each image (namely, one circle) were intercepted within the time interval from −200 to 1,000 ms. The waveforms of ERPs and SSVEPs corresponding to target and non-target images were obtained using the averaged EEG data within the time interval of (−200 1,000) ms.

Target Classification

Single-circle EEG data were firstly processed by spatial filtering methods, and then the target detection was realized by classification algorithms. Four spatial filtering methods, namely, common spatial pattern (CSP), SIgnal-to-noise ratio Maximizer (SIM), task-related component analysis (TRCA), and principal component analysis (PCA) whitening, were compared in this study. The effects of the number of components (from 1 to 50) of different spatial filtering methods on the classification performance were compared. The performance of spatial filtering was evaluated by the followed classification results of the classical Hierarchical Discriminant Component Analysis (HDCA) algorithm, which was adopted as a baseline measure of classification performance for single-circle EEG between target and non-target images (Gerson et al., 2006; Sajda et al., 2010). As a classical classification method widely used in RSVP-BCIs, HDCA algorithm realizes target images recognition based on spatial and temporal projection features of ERP signals. EEG data were firstly divided into 100-ms data segments, and then the feature extraction and classification were conducted according to the spatial and temporal characteristics of the data segments.

To evaluate the performance of the classification methods, four classification algorithms, namely, Support Vector Machine (SVM), Spatially Weighted Fisher linear discriminant (FLD)-PCA (SWFP), Discriminative Canonical Pattern Matching (DCPM), and HDCA, were compared based on this dataset.

The EEG data used for single-circle classification were the data in the time interval of [0, t] ms, "t" might be 200, 300,. . ., 1,000 ms. SIM algorithm was used as a basic spatial filtering method before the performance comparison of the four classification algorithms.

Performance Evaluation

R-square values for each time point were used to show the separability between target and non-target stimuli. For each subject, we selected all the target data and the same amount of non-target data randomly selected to calculate *r*-square values. For each time point, the input was composed of two one-dimensional vectors, which were composed of target data and non-target data, respectively. The *r*-square values of each subject were calculated, and the *r*-square values of all subjects were averaged to obtain the final results, as shown in **Figure 3B**.

Classification performance of single-circle EEG data for target and non-target circles was measured using the area under the receiver operating characteristic (ROC) curve (Fawcett, 2006). ROC curves are used when applications have an unbalanced class distribution, which is typically the case with RSVP-BCI, where the number of target stimulus is much smaller than that of non-target stimuli.

Statistical Analysis

Statistical analyses were conducted using SPSS software (IBM SPSS Statistics, IBM Corporation). One-way repeated-measures analysis of variance (ANOVA) was used to test the difference in the classification performances among different algorithms. The Greenhouse–Geisser correction was applied if the data did not conform to the sphericity assumption by Mauchly's test of sphericity. All pairwise comparisons were Bonferroni corrected. Statistical significance was defined as $p < 0.05$.

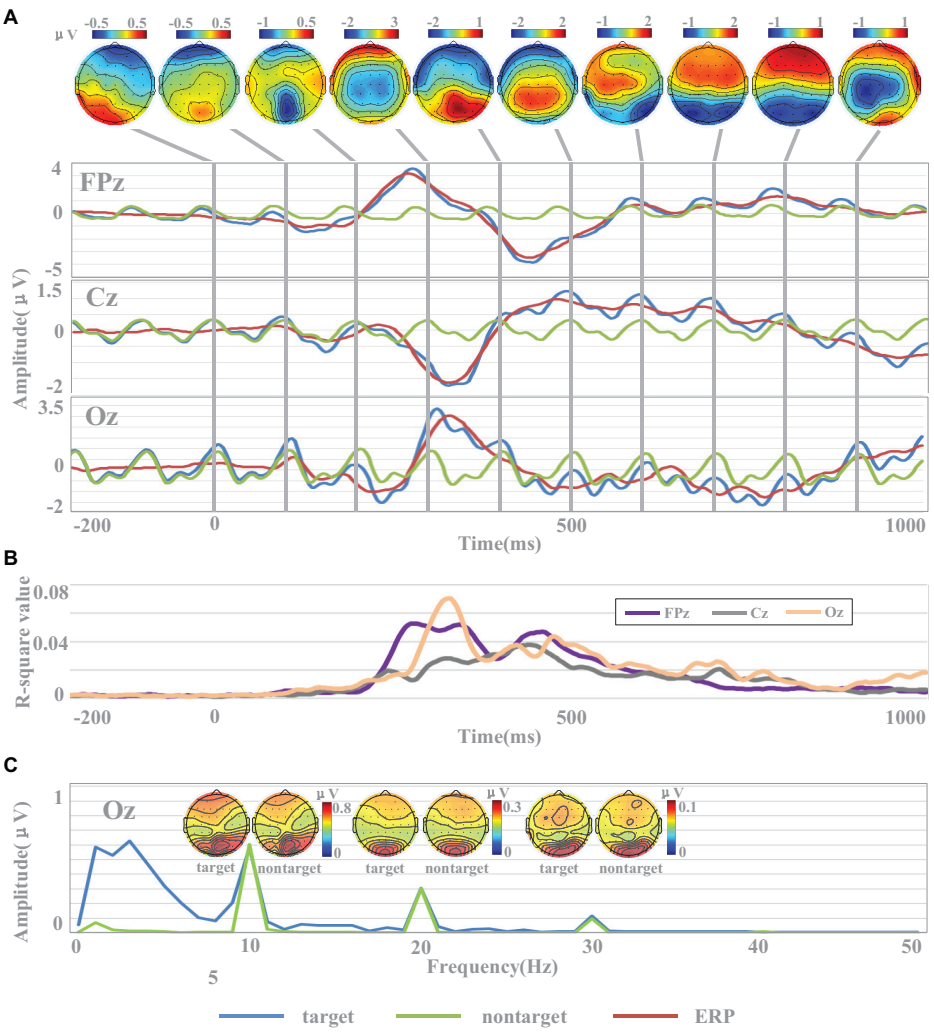


FIGURE 3 | Waveforms and amplitude spectrum of EEG. **(A)** The temporal waveforms of EEG [including target images, non-target images, and event-related potential (ERP) data] and the scalp topographies of amplitudes of ERP. **(B)** The *r*-square values in channels of FPz, Cz, and Oz. **(C)** Spectral characteristics of EEG for target and non-target images.

DATA RECORDING

EEG Data

The dataset is freely available at http://bci.med.tsinghua.edu.cn/download.html. The dataset was the raw continuous data without any processing. It contains 128 MATLAB MAT files corresponding to data from all 64 subjects (approximately 15 GB in total). Data were stored as double-precision floating-point values in MATLAB. Each MAT file covers a group of EEG data. There are two sets of EEG data (groups A and B) for subjects from 1 to 64. The files were named as subject and group indices (i.e., sub1A.mat, sub1B.mat,. . ., sub64A.mat, sub64B.mat). For each file, the data loaded in MATLAB generate two 2-D matrices named "EEGdata1" (block1) and "EEGdata2" (block2) with dimensions of [64, L] (the two dimensions indicate "Electrode index," "Time points," respectively) and two 2-D matrices named "class_labels" and "trigger_positions" with dimensions of [2, 4000]. The parameter of L (the length of time points) might be different for different blocks. The two dimensions indicate "class labels," in which "2" and "1" indicate "non-target images" and "target images," respectively. Each circle corresponds to the EEG data of a visual stimulus image. For each group, the data matrix consists of 8,000 circles (100 circles × 40 trials × 2 blocks), and each circle consists of 64 channels of EEG data. A "Readme.txt" file explains the data structure and other task-related information.

Electrode Position

The electrode positions were listed in a "64-channels.loc" file, which contained all channel locations in polar coordinates. Information for each electrode contained four columns: "Electrode Index," "Degree," "Radius," and "Label." For example, information on the first electrode was as follows: ("1," "−18," "0.51111," and "FP1"), which indicated that the degree is −18, and the radius is 0.51111 for the first electrode (FP1). The electrode file can be used for topographic analysis by the topoplot() function in the EEGLAB toolbox (Delorme and Makeig, 2004).

TECHNICAL VALIDATIONS

Temporal Waveform and Amplitude Spectrum Analysis

To evaluate the signal quality of the dataset, this study analyzed temporal waveform and amplitude spectrum of EEG across all subjects. EEG data were re-referenced to the average of all electrodes. **Figure 3A** shows the temporal waveform of averaged EEG across all subjects. Three representative midline electrodes (FPz, Cz, and Oz) were selected for temporal waveforms display. For each subject, all EEG data corresponding to target and non-target images were averaged. Then, the averaged target and non-target EEG data for each subject were averaged across all subjects. Finally, the cross-subject averaged EEG data corresponding to the non-target images were subtracted from that of the target images to generate the target-related ERP, as shown in **Figure 3A**. To better observe the temporal characters of the SSVEPs, the data were band-pass filtered between 2 and 30 Hz within the time window from −200 to 1,000 ms.

The EEG signals in this dataset were sensitive to target and non-target image stimuli, and the difference of the evoked EEG between the target and non-target image stimuli could be reflected by the ERP components within a short data length at specific brain regions. **Figure 3A** showed the temporal waveforms of EEG for target images, non-target images, and target-related ERP data. The waveform for non-target EEG is a near-sinusoidal signal at 10 Hz with the characteristics of SSVEP. The frequency and phase of the SSVEPs are stable over the 1.2-s stimulation time. The waveforms of ERP located at FPz and Oz showed obvious P300 (FPz: 3.18 μV, Oz: 2.54 μV) and N400 (FPz: −3.49 μV, Oz: −1.29 μV) components. Obviously, the latencies of P300 and N400 components in the prefrontal cortex were significantly smaller than those in the occipital cortex. For example, the latencies of the P300 component in FPz and Oz were 272 and 336 ms, while the latencies of the N400 component were 448 and 484 ms, respectively. While the ERP signal at Cz showed an obvious negative peak appeared around 300 ms (latency: 328 ms, amplitude: −1.29 μV). From the scalp topographies of amplitudes of ERP in **Figure 3A**, it could be found that the areas highly sensitive to ERP response were mainly located in the occipital region and the prefrontal region. For example, these two regions showed significant positive potentials at 300 ms and negative potentials at 400 and 500 ms. The sensitivity of ERPs for the electrode in the parietal region to the stimulation of target images was limited partly because the electrodes were close to the reference electrode.

The results of *r*-square values indicated the separability between target and non-target stimuli, as shown in **Figure 3B**. *R*-square values indicate the importance of features, and the larger the value, the greater the contribution to classification. In the time range of 0–200 ms, the *r*-square values of the three channels were close to 0, which indicated that the features did not contain information valid for classification. After the time of 200 ms, the *r*-square values of the three channels significantly increased, which was consistent with the emergence of the main components of ERP. For example, the *r*-square value of Oz reached the maximum value (0.07) at 340 ms, and at the same time, the ERP of Oz also reached the peak value (2.54 μV). Similar results were also found in Cz and Oz. These results indicated that the emergence of the main components of ERP was accompanied by a greater separability between target and non-target stimuli, and ERP was a potentially effective classification feature. Compared with Cz, the *r*-square values of FPz and Oz were larger, indicating that FPz and Oz contained more effective information and contributed more to classification.

The results of **Figure 3** indicate that the rapid periodic stimulation in RSVP produces a brain response characterized by a "quasi-sinusoidal" waveform whose frequency components are constant in amplitude and phases. **Figure 3C** illustrates the amplitude spectra of EEG evoked by target and non-target images. EEG data were firstly averaged across all subjects, and then the spectrums were calculated by Fast Fourier Transform (FFT) method. As temporal waveforms in **Figure 3A** have shown the non-target EEG as a quasi-sinusoidal signal with stable

frequency and phase, amplitude peaks of EEG at Oz can be observed at 10 Hz and its harmonic frequencies (i.e., 20, 30 Hz) from the frequency information in **Figure 3C**. The amplitudes of fundamental and harmonic components decreased sharply as the response frequency increased (fundamental: 0.60 μV, second harmonic: 0.30 μV, third harmonic: 0.10 μV). Since the signals were filtered from 2 to 30 Hz, the amplitudes in the frequencies above the fourth harmonic were closed to 0. **Figure 3C** also illustrates the scalp topographies of amplitude of target and non-target SSVEP at 10 Hz and its harmonic frequencies. Consistent with previous studies (Gao et al., 2014; Chen et al., 2015a), the occipital area shows the highest amplitude of SSVEPs. In addition to the occipital area, lower amplitude can also be observed at the prefrontal area for components related to stimulus frequency (at 10 and 20 Hz). These characters show very robust and reliable frequency features for the fundamental and harmonic SSVEP components in the dataset and suggest that the RSVP stimulation at 10 Hz in this dataset was stable and reliable.

As the phase characteristic of SSVEPs is synchronous (**Figure 3A**) and the amplitude characteristic is approximate (**Figure 3C**) between target and non-target EEG, target-related ERP signal can be extracted by subtracting non-target EEG from target EEG signals. There were obvious similarities and differences between EEG signals evoked by target images and non-target images in frequency domain. The EEG signals of target images have similar amplitudes of EEG components at the fundamental and harmonic frequencies (fundamental: 0.58 μV, second harmonic: 0.31 μV, third harmonic: 0.11 μV) with that of non-target images. Furthermore, the EEG of the target images contained more powerful low-frequency components (<10 Hz), which were related to ERP. This character suggests that the spectral characteristics provide useful information for the detection of target images.

Evaluating the Performance of Spatial Filtering Methods

Spatial filtering aims to remove signal noise and extract task-related brain activities by using the spatial correlation information of EEG and is frequently applied as a preprocessing method. It has been widely used in EEG-based BCIs. **Figure 4** indicated the performance of different spatial filtering methods in the target/non-target classification task based on the HDCA classification algorithm. Four filtering methods were used to enhance classification performance: CSP, SIM, TRCA, and PCA whitening. CSP consists of finding an optimum spatial filter to maximize the variance difference between two groups of EEG, so as to obtain effective feature vectors for classification (Lotte and Guan, 2011). The algorithm of SIM can be intuitively interpreted as maximizing the signal-to-noise ratio (SNR) in the source space and is an effective tool for spatiotemporal analysis of ERPs (Wu and Gao, 2011). TRCA is the method that extracts task-related components efficiently by maximizing the reproducibility during the task period and can be applied to enhance SNRs of time-locked EEG components such as ERPs (Nakanishi et al., 2018). PCA whitening is a simple and standard procedure to reduce dimension of the data, and it can reduce the complexity by reducing the number of parameters to be estimated (Hyvarinen and Oja, 2000).

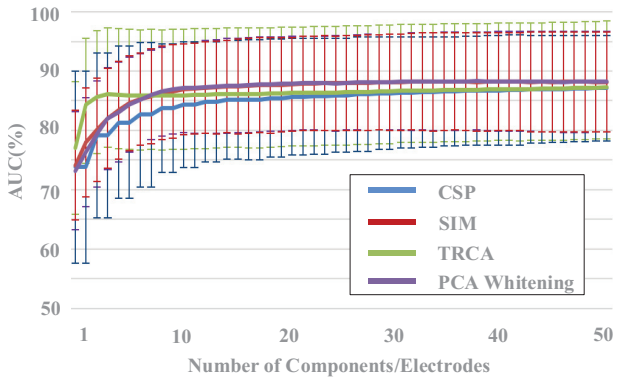


FIGURE 4 | The effect of components number on classification performance (block 1 for training, block 2 for testing).

After the spatial filtering processes, we adopted the HDCA method, which has been widely used in EEG target image detection based on RSVP paradigm, to classify the target and non-target images. For the two blocks in each group of the dataset, EEG data in block 1 were used for training (i.e., to determine parameters of the algorithms), and EEG data in block 2 were used for testing. In addition, both block 1 and block 2 were used for a 10-fold cross-validation to further evaluate the classification performance. Data from all the 62 electrodes were used as the input to the feature extraction and classification analysis. EEG data were firstly divided into 100-ms data segments. Then the feature extraction and classification were conducted according to the spatial and temporal characteristics of the data segments.

The effect of components number of the four spatial filtering methods on classification performance was evaluated. The data length was 400 ms [time window (0 400) ms]. By setting the number of components in the spatial filtering methods (from 1 to 50), the variation of classification performance with the number of components can be obtained (**Figure 4**). The classification performance increased as the number of components increased, especially when the components number was less than 10. For example, the area under the curve (AUC) results for the SIM method were 74.1% ± 9.2%, 78.0% ± 9.2%, 80.0% ± 8.7%, 82.0% ± 8.4%, 83.4% ± 8.3%, 84.6% ± 8.0%, 85.3% ± 7.8%, 85.7% ± 8.0%, 86.3% ± 7.9%, 86.6% ± 7.9% for the components number from 1 to 10, respectively. Especially in the case a small number of components, the TRCA algorithm had the best classification performance. For example, the AUC of TRCA was 77.0, 84.3, and 85.6 as the components number from 1 to 3, respectively, which is far larger than other methods. When the number of components is more than 10, the classification performance no longer changes significantly for all the four methods, and the methods of SIM and PCA whitening show the best performance (SIM: 87.9%, PCA whitening: 88.0%).

A one-way repeated-measures ANOVA showed that there was a statistically significant difference in accuracies among

the four spatial filtering methods for the component numbers of 1 [$F(2.110,268.008) = 4.648$, $p = 0.009$] and from 2 to 50 ($p < 0.001$). Pairwise comparisons showed that the classification accuracies of TRCA were significantly higher ($p < 0.05$) than that of CSP for the component numbers from 2 to 50 and were significantly higher than that of SIM and PCA whitening for the component numbers from 1 to 6. The classification accuracies of SIM and PCA whitening were significantly higher ($p < 0.05$) than that of CSP for the component numbers from 6 to 50 and were significantly higher than that of TRCA for the component numbers from 11 to 50.

Figure 5 shows the results of classification performance for the four spatial filtering methods with different data lengths of EEG. The number of components for the four spatial filtering methods was set to 30. Two validation methods were used, that is, block 1 for training and block 2 for testing (**Figure 5A**) and a 10-fold cross-validation using both blocks 1 and 2 (**Figure 5B**). For each spatial filtering method, the classification accuracy increased obviously as the data length increased when it was less than 500 ms. For example, in **Figure 5A**, the average results of SIM for all subjects were 67.7% ± 7.3%, 80.5% ± 8.8%, 88.1% ± 8.2%, and 91.1% ± 7.2% with the data length from 200 to 500 ms, respectively. The changes of accuracy results were no longer significant when the length of EEG data increased to 600 ms and above.

In addition, there was a significant difference in the classification performance among the different spatial filtering methods. The CSP method corresponded to the worst classification performance, followed by the TRCA method. SIM and PCA whitening methods had higher classification performance with no statistically significant difference. For example, in **Figure 5A**, the classification results were 77.2% ± 10.1%, 78.3% ± 9.4%, 80.5% ± 8.8%, and 80.7% ± 8.8% for the data length of 300 ms in the conditions of CSP, TRCA, SIM, and PCA whitening, respectively. The statistical difference among CSP, SIM, and TRCA was no longer significant when the data length was more than 500 ms. Meanwhile, the high classification results based on EEG with short data lengths indicated that the dataset was collected in a well-designed experimental environment, and the collected EEG data were of high quality.

The 10-fold cross-validation method showed similar results to the original verification method by blocks, i.e., SIM and PCA whitening performed best among the four spatial filtering methods, and HDCA was the best among the four classification methods. The difference between the two validation methods was that the accuracies and variances of the 10-fold cross-validation method were slightly higher and smaller than the method by blocks, respectively. For example, the classification results for CSP, SIM, TRCA, and PCA whitening were 66.0% ± 7.0%, 70.2% ± 7.1%, 67.9% ± 7.0%, and 70.3% ± 7.1% and 63.4% ± 7.1%, 67.7% ± 7.3%, 65.4% ± 7.3%, and 67.8% ± 7.3% for 10-fold cross-validation method and validation method by blocks, respectively. This was due to the fact that the 10-fold cross-validation method used more data for training than the original verification method by blocks. Since the two validation methods have shown similar results, we only chose the classification results of the validation method by blocks to perform the statistical analysis in this study.

A one-way repeated-measures ANOVA showed that there was a statistically significant difference in accuracies among the four spatial filtering methods for the data length of 200 ms [$F(1.326,168.403) = 76.929$, $p < 0.001$], 300 ms [$F(1.324,168.179) = 115.527$, $p < 0.001$], 400 ms [$F(1.204,152.967) = 128.453$, $p < 0.001$], 500 ms [$F(1.333,169.256) = 124.089$, $p < 0.001$], 600 ms [$F(1.247, 58.402) = 131.426$, $p < 0.001$], 700 ms [$F(1.248,158.528) = 101.262$, $p < 0.001$], 800 ms [$F(1.409,178.955) = 100.214$, $p < 0.001$], 900 ms [$F(1.404,178.285) = 99.643$, $p < 0.001$], and 1,000 ms [$F(1.350,171.387) = 102.250$, $p < 0.001$]. Pairwise comparisons showed that the classification accuracies of SIM and PCA whitening were significantly higher ($p < 0.001$) than those of CSP and TRCA for the data length from 200 to 1,000 ms. The classification accuracies of TRCA were significantly higher ($p < 0.01$) than that of CSP for the data length from 200 to 300 ms and were significantly lower ($p < 0.001$) than that of CSP for the data length from 400 to 1,000 ms. There was no significant

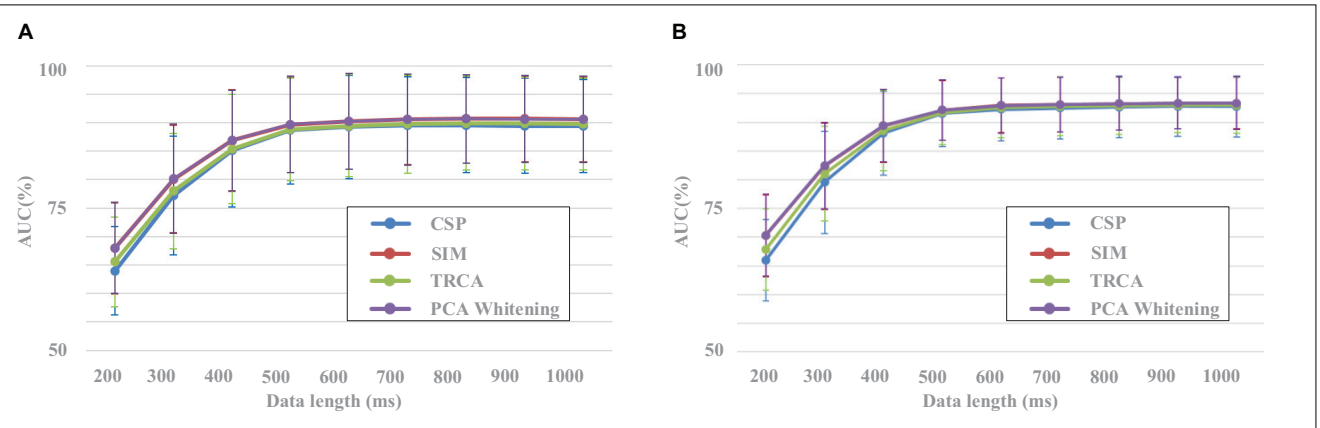


FIGURE 5 | Performance of different data lengths for spatial filtering methods **(A)** Block 1 for training and block 2 for testing. **(B)** Result of 10-fold cross-validation using both blocks 1 and 2.

difference between SIM and PCA whitening for the performance of classification.

Evaluating the Performance of Classification Methods

In addition to the evaluation of spatial filtering methods, the dataset can also be used to evaluate the performance of classification methods. **Figure 6** indicated the performance of different classification methods with the EEG data length from 200 to 1,000 ms. After preprocessing with the SIM method, EEG data for each image were classified by four different algorithms including SVM, SWFP, DCPM, and HDCA. SVM finds a separating hyper-plane that maximizes the margin between the two classes. SWFP is based on a two-step linear classification of event-related responses using FLD classifier and PCA for dimensionality reduction (Alpert et al., 2014). DCPM performs well in classifying the miniature AVePs by first suppressing the common-mode noise of the background EEG and then recognizing canonical patterns of ERPs (Xiao et al., 2020). Two validation methods were used, that is, block 1 for training and block 2 for testing (**Figure 6A**), and a 10-fold cross-validation using both blocks 1 and 2 (**Figure 6B**). As shown in **Figure 6A**, HDCA had the best classification performance, while the other three algorithms had approximately a similar classification performance. This was especially true when the data length was less than 500 ms. For example, the AUC results for HDCA were 67.7% ± 7.3%, 80.5% ± 8.8%, 88.1% ± 8.2%, and 91.1% ± 7.2% for single-circle EEG classification between target and non-target images with the data length of 200, 300, 400, and 500 ms, respectively. When the data length is greater than 500 ms, the performance of the four classification algorithms is similar, while the classification performance of the HDCA algorithm is still the best. **Figure 6B** indicated the similar results as **Figure 6A**, and the only difference was that the SVM method performed the worst.

A one-way repeated-measures ANOVA based on the validation method by blocks showed that there was a statistically significant difference in accuracies among the four classification methods for the data length of 200 ms [$F(2.124,269.799) = 144.651$, $p < 0.001$], 300 ms [$F(1.942,246.670) = 55.645$, $p < 0.001$], 400 ms [$F(2.095,266.046) = 42.243$, $p < 0.001$], 500 ms [$F(2.183,277.251) = 38.436$, $p < 0.001$], 600 ms [$F(2.362,299.935) = 35.408$, $p < 0.001$], 700 ms [$F(3,381) = 27.146$, $p < 0.001$], 800 ms [$F(2.820,358.107) = 33.019$, $p < 0.001$], 900 ms [$F(2.601,330.287) = 29.985$, $p < 0.001$], and 1,000 ms [$F(3,381) = 32.344$, $p < 0.001$]. Pairwise comparisons showed that the classification accuracies of HDCA were significantly higher ($p < 0.001$) than that of SVM, SWFP, and DCPM for the data length from 200 to 1,000 ms. The classification accuracies of SVM were significantly higher ($p < 0.05$) than that of SWFP for the data length from 400 to 1,000 ms and were significantly higher ($p < 0.05$) than that of DCPM for the data length from 900 to 1,000 ms. The classification accuracies of DCPM were significantly higher ($p < 0.01$) than that of SWFP for the data length from 300 to 1,000 ms.

Evaluating the Performance of Cross-Subject Zero-Training Methods

The dataset can be used to study zero-training classification methods of RSVP-based BCIs. To improve the performance of the system, most of the current RSVP-based BCIs adopt supervised feature extraction and classification algorithms that require system calibration. The long time in training data collection and algorithm template extraction processes bring challenges to system practicability and user experience. With benefits from the large scale of the dataset that contains a total of 64 subjects, 10,240 trials, 1,024,000 image circles, and 102,400 s of 64-channel EEG data, it is possible to extract common information of EEG for target classification. A cross-subject strategy can be used to design zero-training algorithms suitable for target identification in the RSVP paradigm.

In this paper, the dataset was used to design a zero-training classification algorithm based on a cross-subject template. The performance was estimated using a leave-one-subject-out cross-validation. EEG data of each subject were trained separately to obtain his or her algorithm template parameters for the HDCA algorithm. In the testing session, by using cross-subject template, the EEG classification performance of one subject was

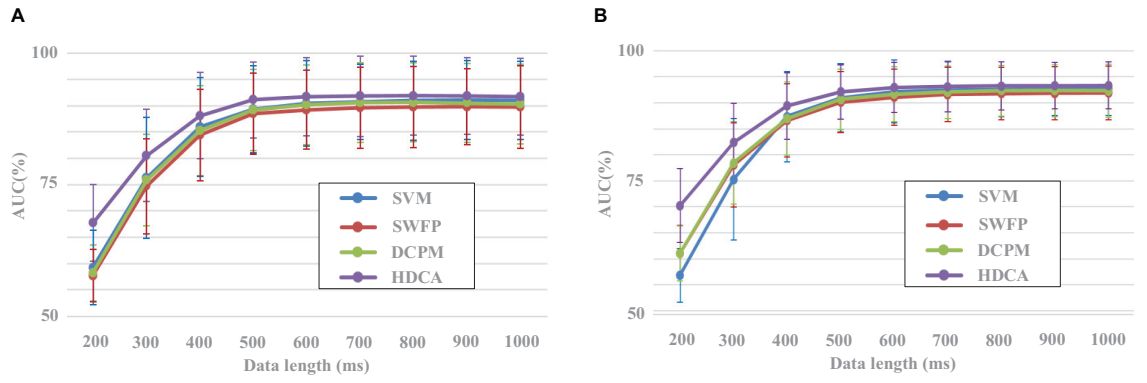


FIGURE 6 | Performance of different classification methods with different data lengths. **(A)** Block 1 for training and block 2 for testing. **(B)** Result of 10-fold cross-validation using both blocks 1 and 2.

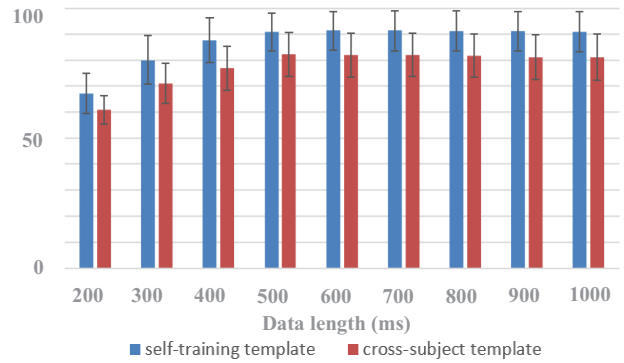


FIGURE 7 | Performance of cross-subject zero-training and self-training methods (block 1 for training and block 2 for testing).

determined by the voting results of the algorithm templates of the other 63 subjects, all of whom had an equal voting weight. **Figure 7** showed the performance of cross-subject zero-training method using the HDCA algorithm. Pairwise comparisons showed that the classification accuracies of the traditional self-training method were significantly higher than that of the cross-subject method for the data length of 200 ms [$F(1,127) = 83.0101$, $p < 0.001$], 300 ms [$F(1,127) = 164.440$, $p < 0.001$], 400 ms [$F(1,127) = 195.524$, $p < 0.001$], 500 ms [$F(1,127) = 137.263$, $p < 0.001$], 600 ms [$F(1,127) = 143.973$, $p < 0.001$], 700 ms [$F(1,127) = 139.003$, $p < 0.001$], 800 ms [$F(1,127) = 139.555$, $p < 0.001$], 900 ms [$F(1,127) = 151.889$, $p < 0.001$], and 1,000 ms [$F(1,127) = 141.892$, $p < 0.001$]. Although the performance of cross-subject method was lower than the traditional self-training method, it still achieved good performance for more than 80% of AUC when the data length was more than 400 ms. For example, the AUCs were 82.2% ± 8.4% and 90.8% ± 7.4% by using cross-subject and self-training templates, respectively, when the data length was 500 ms. The results indicated that a variety of cross-subject information could be mined from the dataset. By using the dataset appropriately, we can effectively design algorithms that do not require system calibration. With the mining of more effective information contained in the dataset, it is believed that the performance of zero-training algorithm can be further improved and even closer to the performance of the training methods. This dataset provides sufficient data for the development of zero-training algorithms that can promote the practical application of RSVP-based BCIs.

DISCUSSION AND CONCLUSION

This study presents a benchmark dataset for studying RSVP-based BCIs. Distinct ERP and SSVEP features in temporal, frequency, and spatial domains prove the high quality of data. The examples on evaluating classification performance further demonstrate high efficiency of the dataset for evaluating methods in target image detection.

In this study, continuous image stimulation was divided into periodic segments to resist fatigue and ensure the high quality and reliability of EEG signals. To reduce the interference of blinking on EEG, subjects were instructed to blink between trials rather than within the image sequence of stimuli, and they initiated the next trial by pressing a button. At the same time, subjects were given enough rest between blocks until they felt comfortable to start the next block. In this study, no strict experimental interruption time was set, which fully guaranteed the quality of EEG signals. The impact of rest time can be considered in future practical applications.

Besides the above technical validations proposed in this study, the dataset can be further analyzed in a variety of different ways. In fact, although remarkable progresses have been made in RSVP-BCI, there are still many defects to be solved. Firstly, the parameters of RSVP-BCI need to be optimized to meet different application requirements; secondly, the characteristics of SSVEP and ERP that are evoked by the RSVP paradigm require further investigation; thirdly, the separation methods of SSVEP and ERP are not effective. This dataset can be used for developing methods to address these limitations. On one hand, the dataset can be used to design system diagrams toward different applications. The optimization of parameters is very important for the design and implementation of a practical BCI system (Zhang and Gao, 2019; Lees et al., 2020). For example, the effect of time interval between target images on EEG characteristics can not only inspire the design of optimal RSVP stimulation paradigm but also deepen the understanding of attentional blink. Regarding the phase of the EEG, although other experimental paradigms such as SSVEP-BCIs have already shown indicators of phase character of evoked EEG such as latency, very few studies based on RSVP-BCIs explored phase characters. The evoked EEG phase in the RSVP paradigm must contain higher cognitive mechanisms, which makes the relevant research more significant. Besides, the number of electrodes and electrode locations can be optimized using the 64-channel dataset. On the other hand, the dataset can be used to develop computational models for ERPs and SSVEPs. The high SNR of ERP and SSVEPs from the dataset could be helpful for exploring the intrinsic properties of ERP and SSVEP harmonics. For example, the way to characterize the phases of the fundamental and harmonic SSVEP components still remains unknown. Furthermore, it is of great scientific significance to study the methods for separating ERP and SSVEP signals and the temporal dynamics and phase relations between them. The problem has not been well solved so far, and this dataset provides rich resources for the related studies.

In future work, the dataset can be improved in the following directions. First, data evoked by stimulus images with different frequencies will be included. In this study, the stimulation frequency was set to the most commonly used 10 Hz. EEG data with different frequencies may help to reveal the effect of workload on EEG. Secondly, more types of target sparsity will be included. As the target sparsity is set as 1∼4% in this study, the probability of target images can be further increased to verify the relationship between target density and the EEG signals. Third, data records from the same group of subjects on different days will be provided for developing the session-to-session transfer approach (Zhao et al., 2019), which can facilitate system calibration in an online BCI.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found at: http://bci.med.tsinghua.edu.cn/download.html.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Research Ethics Committee of Tsinghua University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YW, LZ, and XG designed the research. SZ performed the research. SZ and YW analyzed the data. All authors wrote the manuscript and contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Key R&D Program of Guangdong Province, China, under grant 2018B030339001, the National Key R&D Program of China under grant 2017YFB1002505, the Strategic Priority Research Program of Chinese Academy of Sciences under grant XDB32040200, the Fundamental Research Funds for the Central Universities under Grant FRF-TP-20-017A1, the Fundamental Research Funds for the University of Science and Technology Beijing under Grant FRF-BD-19-012A, and the National Natural Science Foundation of China under grant 61431007.

ACKNOWLEDGMENTS

We would like to thank the subjects who participated in this study, D. Zhang, H. Zhao, and L. Zheng for their assistance in data collection, and C. Yang and H. Wu for their work in making the web page for data download.

REFERENCES

Abibullaev, B., and Zollanvari, A. (2019). *Event-Related Potentials (P300, EEG) - BCI dataset. IEEE DataPort*. Kazakhstan: Nazarbayev University, doi: 10.21227/8aae-d579

Alpert, G. F., Manor, R., Spanier, A. B., Deouell, L. Y., and Geva, A. B. (2014). Spatiotemporal representations of rapid visual target detection: a single-trial EEG classification algorithm. *IEEE Trans. Biomed. Eng.* 61, 2290–2303. doi: 10.1109/TBME.2013.2289898

Bigdely-Shamlo, N., Vankov, A., Ramirez, R. R., and Makeig, S. (2008). Brain activity-based image classification from rapid serial visual presentation. *IEEE Trans. Neural Syst. Rehabil. Eng.* 16, 432–441. doi: 10.1109/TNSRE.2008.2003381

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision.* 10, 433–436. doi: 10.1163/156856897X00357

Chen, X. G., Wang, Y. J., Gao, S. K., Jung, T. P., and Gao, X. R. (2015a). Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain-computer interface. *J. Neural Eng.* 12:046008. doi: 10.1088/1741-2560/12/4/046008

Chen, X. G., Wang, Y. J., Nakanishi, M., Gao, X. G., Jung, T. P., and Gao, S. K. (2015b). High speed spelling with a noninvasive brain-computer interface. *Proc. Natl. Acad. Sci. U.S.A.* 112, E6058–E6067. doi: 10.1073/pnas.1508080112

Cho, H., Ahn, M., Ahn, S., Kwon, M., and Jun, S. C. (2017). EEG datasets for motor imagery brain-computer interface. *Gigascience* 6:8. doi: 10.1093/gigascience/gix034

Cohen, M. (2014). *Analyzing Neural Time Series Data: Theory and Practice*. Cambridge, MA: The MIT Press.

Delorme, A., and Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods.* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recogn. Lett.* 27, 861–874. doi: 10.1016/j.patrec.2005.10.010

Gao, S. K., Wang, Y. J., Gao, X. R., and Hong, B. (2014). Visual and auditory brain-computer interfaces. *IEEE Trans. Biomed. Eng.* 61, 1436–1447. doi: 10.1109/TBME.2014.2300164

Gerson, A. D., Parra, L. C., and Sajda, P. (2006). Cortically coupled computer vision for rapid image search. *IEEE Trans. Neural Syst. Rehabit. Eng.* 14, 174–179. doi: 10.1109/TNSRE.2006.875550

Han, C. H., Muller, K. R., and Hwang, H. J. (2020). Brain-switches for asynchronous brain-computer interfaces: a systematic review. *Electronics* 9:422. doi: 10.3390/electronics9030422

Huang, L. T., Zhao, Y. Q., Zeng, Y., and Lin, Z. M. (2017). BHCR: RSVP target retrieval BCI framework coupling with CNN by a Bayesian method. *Neurocomputing* 238, 255–268. doi: 10.1016/j.neucom.2017.01.061

Hyvarinen, A., and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Netw.* 13, 411–430. doi: 10.1016/s0893-6080(00)00026-5

Kaya, M., Binli, M. K., Ozbay, E., Yanar, H., and Mishchenko, Y. (2018). A large electroencephalographic motor imagery dataset for electroencephalographic brain computer interfaces. *Sci. Data* 5:180211. doi: 10.1038/sdata.2018.211

Lee, M. H., Kwon, O. Y., Kim, Y. J., Kim, H. K., Lee, Y. E., Williamson, J., et al. (2019). EEG dataset and OpenBMI toolbox for three BCI paradigms: an investigation into BCI illiteracy. *Gigascience* 8, 1–16. doi: 10.1093/gigascience/giz002

Lees, S., Dayan, N., Cecotti, H., Mccullagh, P., Maguire, L., Lotte, F., et al. (2017). A review of rapid serial visual presentation-based brain-computer interfaces. *J. Neural Eng.* 15:021001. doi: 10.1088/1741-2552/aa9817

Lees, S., McCullagh, P., Payne, P., Maguire, L., Lotte, F., and Coyle, D. (2020). Speed of rapid serial visual presentation of pictures, numbers and words affects event-related potential-based detection accuracy. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28:122. doi: 10.1109/TNSRE.2019.2953975

Lioi, G., Cury, C., Perronnet, L., Mano, M., Bannier, E., Lecuyer, A., et al. (2019). Simultaneous MRI-EEG during a motor imagery neurofeedback task: an open access brain imaging dataset for multi-modal data integration. *bioRxiv[Preprint]*. doi: 10.1101/862375

Lotte, F., and Guan, C. T. (2011). Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms. *IEEE Trans. Biomed. Eng.* 58, 355–362. doi: 10.1109/TBME.2010.2082539

Manor, R., Mishali, L., and Geva, A. B. (2016). Multimodal neural network for rapid serial visual presentation brain computer interface. *Front. Comp. Neurosci.* 10:130. doi: 10.3389/fncom.2016.00130

Nakanishi, M., Wang, Y. J., Chen, X. G., Wang, Y. T., Gao, X. R., and Jung, T. P. (2018). Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis. *IEEE Trans. Biomed. Eng.* 65, 104–112. doi: 10.1109/TBME.2017.2694818

Sajda, P., Gerson, A., and Parra, L. (2003). High-throughput image search via single-trial event detection in a rapid serial visual presentation task [M]. *1st International IEEE/EMBS Conference on Neural Engineering. Capri Italy*. 2003, 7–10. doi: 10.1109/CNE.2003.1196297

Sajda, P., Pohlmeyer, E., Wang, J., Parra, L., Christoforou, C., Dmochowski, J., et al. (2010). In a blink of an eye and a switch of a transistor: cortically

coupled computer vision. *Proc. IEEE*. 98, 462–478. doi: 10.1109/JPROC.2009.2038406

Singh, A., and Jotheeswaran, J. (2018). "P300 brain waves instigated semi supervised video surveillance for inclusive security systems. Advances in Brain Inspired Cognitive Systems," in *Proceedings of the 9th International Conference*, China, 184–194. doi: 10.1007/978-3-030-00563-4_18

Stegman, P., Crawford, C. S., Andujar, M., Nijholt, A., and Gilbert, J. E. (2020). Brain-computer interface software: a review and discussion. *IEEE Trans. Hum. Mach. Syst*. 50:115. doi: 10.1109/THMS.2020.2968411

Vaineau, E., Barachant, A., Andreev, A., Rodrigues, P. C., Cattan, G., and Congedo, M. (2019). Brain invaders adaptive versus non-adaptive P300 brain-computer interface dataset. *arXiv[Preprint]*. doi: 10.5281/zenodo.1494163

Wang, Y. J., Chen, X. G., Gao, X. R., and Gao, S. K. (2017). A benchmark dataset for SSVEP-based brain-computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 1746–1752. doi: 10.1109/TNSRE.2016.2627556

Wu, Q. J., Yan, B., Zeng, Y., Zhang, C., and Tong, L. (2018). Anti-deception: reliable EEG-based biometrics with real-time capability from the neural response of face rapid serial visual presentation. *Biomed. Eng.*17:55. doi: 10.1186/s12938-018-0483-7

Wu, W., and Gao, S. K. (2011). "Learning Event-Related Potentials (ERPs) from multichannel EEG recordings: a spatio-temporal modeling framework with a fast estimation algorithm [M]," in *Proceedings of the 33rd Annual International Conference of the IEEE Engineering-in-Medicine-and-Biology-Society (EMBS)*, Boston, MA, 6959–6962.

Xiao, X. L., Xu, M. P., Jin, J., Wang, Y. J., Jung, T. P., and Ming, D. (2020). Discriminative canonical pattern matching for single-trial classification of ERP components. *IEEE Trans. Biomed. Eng.* 67, 2266–2275. doi: 10.1109/TBME.2019.2958641

Zhang, S. G., and Gao, X. R. (2019). The effect of visual stimuli noise and fatigue on steady-state visual evoked potentials. *J. Neural Eng.* 16:056023. doi: 10.1088/1741-2552/ab1f4e

Zhang, S. G., Han, X., Chen, X. G., Wang, Y. J., Gao, S. K., and Gao, X. R. (2018). A study on dynamic model of steady-state visual evoked potentials. *J. Neural Eng.* 15:046010. doi: 10.1088/1741-2552/aabb82

Zhang, S. G., Han, X., and Gao, X. R. (2020). Studying the effect of the pre-stimulation paradigm on steady-state visual evoked potentials with dynamic models based on the zero-pole analytical method. *Tsingh. Sci. Technol.* 25, 435–446. doi: 10.26599/TST.2019.9010028

Zhao, H. Z., Wang, Y. J., Sun, S., Pei, W. H., and Chen, H. D. (2019). "Obviating session-to-session variability in a rapid serial visual presentation-based brain-computer interface,"in *Proceedings of the 9TH International IEEE/EMBS Conference on Neural Engineering (NER)*, San Francis, CA, 174. doi: 10.1109/NER.2019.8716892

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

DATA REPORT
published: 19 October 2020
doi: 10.3389/fnins.2020.566147

Motor Imagery Under Distraction—An Open Access BCI Dataset

Stephanie Brandl[1]* *and Benjamin Blankertz*[2]

[1] Department of Machine Learning, Technische Universität Berlin, Berlin, Germany, [2] Department of Neurotechnology, Technische Universität Berlin, Berlin, Germany

Keywords: brain-computer interface, motor imagery, out-of-lab scenarios, artifacts, steady-state visual evoked potential (SSVEP), vibro-tactile stimulation

OPEN ACCESS

Edited by:
Andrea Kübler,
Julius Maximilian University of Würzburg, Germany

Reviewed by:
Pavel Bobrov,
Institute of Higher Nervous Activity and Neurophysiology (RAS), Russia
Yu Zhang,
Stanford University, United States

***Correspondence:**
Stephanie Brandl
stephanie.brandl@tu-berlin.de

Specialty section:
This article was submitted to Neural Technology, a section of the journal Frontiers in Neuroscience

Received: *27 May 2020*
Accepted: *21 August 2020*
Published: *19 October 2020*

Citation:
Brandl S and Blankertz B (2020) Motor Imagery Under Distraction— An Open Access BCI Dataset. Front. Neurosci. 14:566147. doi: 10.3389/fnins.2020.566147

1. INTRODUCTION

Research studies in the field of Brain-Computer Interfaces (BCI) mostly take place in controlled lab environments. To move BCIs into the real world and everyday life situations it is crucial to bring research out of those controlled environments and into more realistic scenarios.

Recently, various studies have been recorded in classrooms, cars or realistic tugboat simulators (Blankertz et al., 2010; Brouwer et al., 2017; Ko et al., 2017; Miklody et al., 2017). Mobile BCIs even allow participants to move freely during the recording (Lotte et al., 2009; Castermans et al., 2011; De Vos et al., 2014; Wriessnegger et al., 2017; von Lühmann et al., 2019). Other studies have been carried out with paralyzed, locked-in or completely locked-in users or with participants recovering from stroke (Neuper et al., 2003; Ang et al., 2011; Leeb et al., 2013; Höhne et al., 2014; Hwang et al., 2017; Han et al., 2019; Lugo et al., 2020).

However, so far there has not been a BCI study where distractions are investigated systematically. We have recorded a motor imagery-based BCI study ($N = 16$) under five types of distractions that mimic out-of-lab environments and a control task where no distraction was added. The secondary tasks include watching a flickering video, searching the room for a specific number, listening to news, closing the eyes and vibro-tactile stimulation.

Many BCI datasets have been published, e.g., in context of the BNCI Horizon 2020 initiative[1], 4 BCI competitions have had a big impact on the research community (Sajda et al., 2003; Blankertz et al., 2004, 2006; Tangermann et al., 2012) and still datasets are made available (Shin et al., 2016; Cho et al., 2017; Kaya et al., 2018). We want to contribute further by publishing this BCI dataset with multiple distractor conditions. This report provides a summary of the design and experimental setup of the study. We also show group-level results on event-related synchronization and desynchronization, results on a standard classification pipeline and power spectra for all secondary tasks. Apart from the dataset[2], code for the analysis is also publicly available[3] and a more advanced analysis can be found in Brandl et al. (2016).

[1] https://bnci-horizon-2020.eu/database/data-sets
[2] https://depositonce.tu-berlin.de/handle/11303/10934.2
[3] https://github.com/stephaniebrandl/bci-under-distraction

2. METHODS

2.1. Participants

Sixteen participants (six female, average age 26.3 ± 1.9 years) volunteered to participate in this study. Three volunteers had previously participated in another BCI experiment. All instructions were given in German requiring basic language skills. Volunteers were reimbursed for their participation in the study except for three employees of the TU Berlin Machine Learning Group. All participants were instructed on the experimental procedures prior to signing an informed consent. This study was conducted according to the declaration of Helsinki and was approved by the Ethics Committee of the Charite-Universitätsmedizin Berlin (approval number: EA4/012/12).

2.2. Data Acquisition

EEG signals were recorded with a Fast'n Easy Cap (EasyCap GmbH) with 63 wet Ag/AgCl electrodes which were placed at symmetrical positions according to the international 10–20 system (Jasper, 1958) referenced to the nose. We used two 32-channel amplifiers (BrainAmp, BrainProducts) to amplify the signals, which were sampled at 1,000 Hz. Data was recorded in the period of 15 April–18 July 2014 at TU Berlin and raw data without any preprocessing was made publicly available[1].

2.3. Experimental Setup

During the experiment, the participants were sitting in a comfortable armchair at a distance of 1m in front of a 24" computer screen. Auditory instructions were given via headphones.

Each experimental session lasted about 3 h including preparation and about 90 min of signal recording. Before the main experiment, we recorded eight trials in which participants had to alternately keep their eyes open or closed for 15 s.

The main experiment was divided into seven runs à 10 min with 72 trials per run. One trial lasted 4.5 s and was defined by one motor imagery task with an additional secondary task except for the first run. The first run served as a calibration phase without feedback and distraction tasks. The subsequent runs included three blocks à four trials (two left and two right) of each secondary tasks (72 trials per run). The blocks were presented in a random order to minimize sequence effects.

2.3.1. Primary Task

At the beginning of each trial, instructions for left or right hand motor imagination were given over headphones (*links* and *rechts* as the instructions were in German). This was the primary task in this study. At the end of the trial the participant received a *stop* command followed by a break of 2.5 s, after which the next trial started.

Participants were asked to choose one haptic hand movement. Several strategies for motor imagery were presented to the participants to choose from. The majority chose to imagine squeezing a soft ball—other strategies involved opening a water tap, piano playing or using a salt shaker.

Auditory online feedback was given in the six runs after the calibration to keep the motivation up. The online feedback was trained on the calibration data and based on Laplacian filters of the C3 and C4 electrodes (McFarland et al., 1997) and regularized linear discriminant analysis (RLDA, Friedman, 1989). For this, EEG data was downsampled to 100 Hz, Laplacian filters of C3 and C4 were calculated and the data was band-pass filtered in the ranges 9–13 and 18–26 Hz with a Butterworth filter of order 5. Data was then cut into epochs of 750–3,500 ms and an RLDA classifier was trained on the logarithm of variances as features. During the feedback phase, EEG data was downsampled and band-pass filtered as before, projected on the Laplacian filters and the trained classifier applied on the log-variance features. Furthermore, we applied pooled-mean adaptation to continue training the classifier during the feedback phase (Vidaurre et al., 2010). Classification averaged across all participants reached an accuracy of 57.05%. Auditory feedback was given after the *stop* command as *decision left* (*Entscheidung links*) or *decision right* (*Entscheidung rechts*) during the 2.5 s break. Online classification was performed with the BBCI toolbox in MATLAB[4].

2.3.2. Secondary Tasks

We simulated a pseudo-realistic environment by adding six secondary tasks on top of the primary motor imagery task to the experimental setup. They were selected to cover different types of distractions in an out-of-lab scenario.

1. *Clean*
   This condition served as a control task where no additional distraction was added.
2. *Eyes-Closed*
   Participants were asked to close their eyes before the motor imagery trial started and to keep them closed until the trial finished. Here, we expected a power increase in the alpha band (8–12 Hz) due to the closed eyes to overlap with the motor task related mu rhythm (8–13 Hz). This task was also the primary reason for providing all instructions and feedback auditorily instead of visually.
3. *News*
   Short sequences of a public newscast (*Tagesschau*) were played over the headphones with current news (January/February 2014) and news from 1994. Each sequence was only played once in each experiment. We expected the participants to be cognitively distracted and the auditory cortex to be activated during the motor imagery task which might influence the motor imagery performance.
   During the experiment, we did not assess active listening of the participants.
4. *Numbers*
   For this task, 26 sheets of paper with a randomly mixed letter-number combination were set up on the wall in front of the participants and also on the left and right side of the room. This implies that participants needed to turn their head in order to see the sheets. For each trial a new window appeared on the screen asking the participants to search the room for a particular letter to match with a stated number and to read it out loud. Each combination was shown 2–3 times to all

[4]https://github.com/bbci/bbci_public

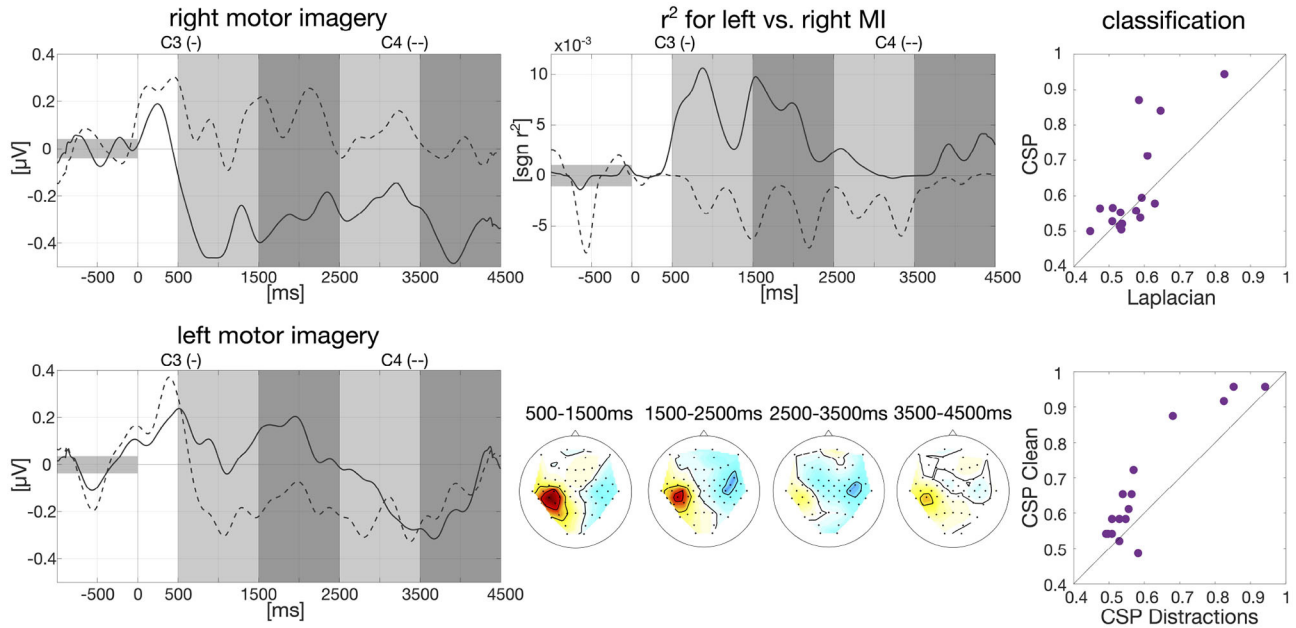


FIGURE 1 | Baseline analysis for calibration data and first classification results. Left: Group-average envelopes of C3 (solid) and C4 (dashed) electrodes for right and left motor imagery trials. Center: Group-average signed r^2 values evolving over time within trials as times series for C3 and C4 (upper) and patterns for four different intervals (lower). Right: Classification results on motor imagery data left vs. right (one dot per participant): comparison offline CSP vs. online Laplacian (upper) and comparison CSP without distractions vs. CSP with distractions (lower).

participants. We counted how often the letters were found. Out of 72 trials, 59.7 combinations were successfully found on average. This task was expected to cause both a high cognitive distraction and additional muscular artifacts.

5. *Flicker*
   A flickering stimulus with alternating gray shades at a frequency of 10 Hz was presented on the screen. We included this task to analyze the influence of the *steady state visually evoked potential* (SSVEP) (Morgan et al., 1996).
6. *Stimulation*
   We placed two coin vibration motors with a diameter of 3 cm on the insides of both forearms, one over each wrist and the other just below the elbows. To investigate the interference of *steady state vibration somatosensory evoked potential* (SSVSEP, Tobimatsu et al., 1999; Brouwer and Van Erp, 2010) on the motor imagery task, vibrotactile stimulation was carried out with carrier frequencies of 50 and 100 Hz, each modulated at 9, 10, and 11 Hz.

2.4. Baseline Analysis

We show group-level results of event-related synchronization and desynchronization (ERS/ERD, see **Figure 1**) which can be observed during motor imagination and execution (Pfurtscheller, 1992). Data analysis was also performed with the BBCI toolbox for MATLAB[4].

Data from the calibration session was band-pass filtered in the frequency band of 9–13 Hz with a 3rd order zero-phase Butterworth filter and cut into epochs for each participant individually, starting 1,000 ms prior to trial onset until 4,500 ms after trial onset. The envelope was then calculated on the group average based on the Hilbert transformation with a moving average window of 200 ms. Baseline correction was applied, i.e., the average EEG amplitude in the interval of 1,000 ms prior to trial onset was subtracted. The resulted smoothed envelope is presented in **Figure 1** for the electrodes C3 and C4. Here, we clearly see desynchronization effects in C3 for right hand motor imagery and C4 for left hand motor imagery starting around 500 ms after trial onset.

We further calculated signed biserial correlation coefficients (r^2) on the smoothed group-average envelope to determine which EEG channels show the most discriminative information for left and right hand motor imagery. Results can be examined in **Figure 1** where the scalp patterns of both left and right motor cortex carry relevant class information especially in the beginning of the trial which matches findings in the literature (Pfurtscheller, 1992). Above the scalp patterns, we show the time course over an average of all epochs of the r^2-values for C3 and C4. Here, we can see that on average 500–2,000 ms after trial onset the two channels carry import information to separate right and left motor imagery as indicated by r^2.

We also conducted an offline classification with Common Spatial Patterns (CSP, Ramoser et al., 2000) in comparison to the online classification with Laplacian filters. Individual frequency bands between 8 and 30 Hz and time intervals between 250 and 4,500 ms after stimulus onset were selected for each participant as described in Blankertz et al. (2007). Data was then band-pass

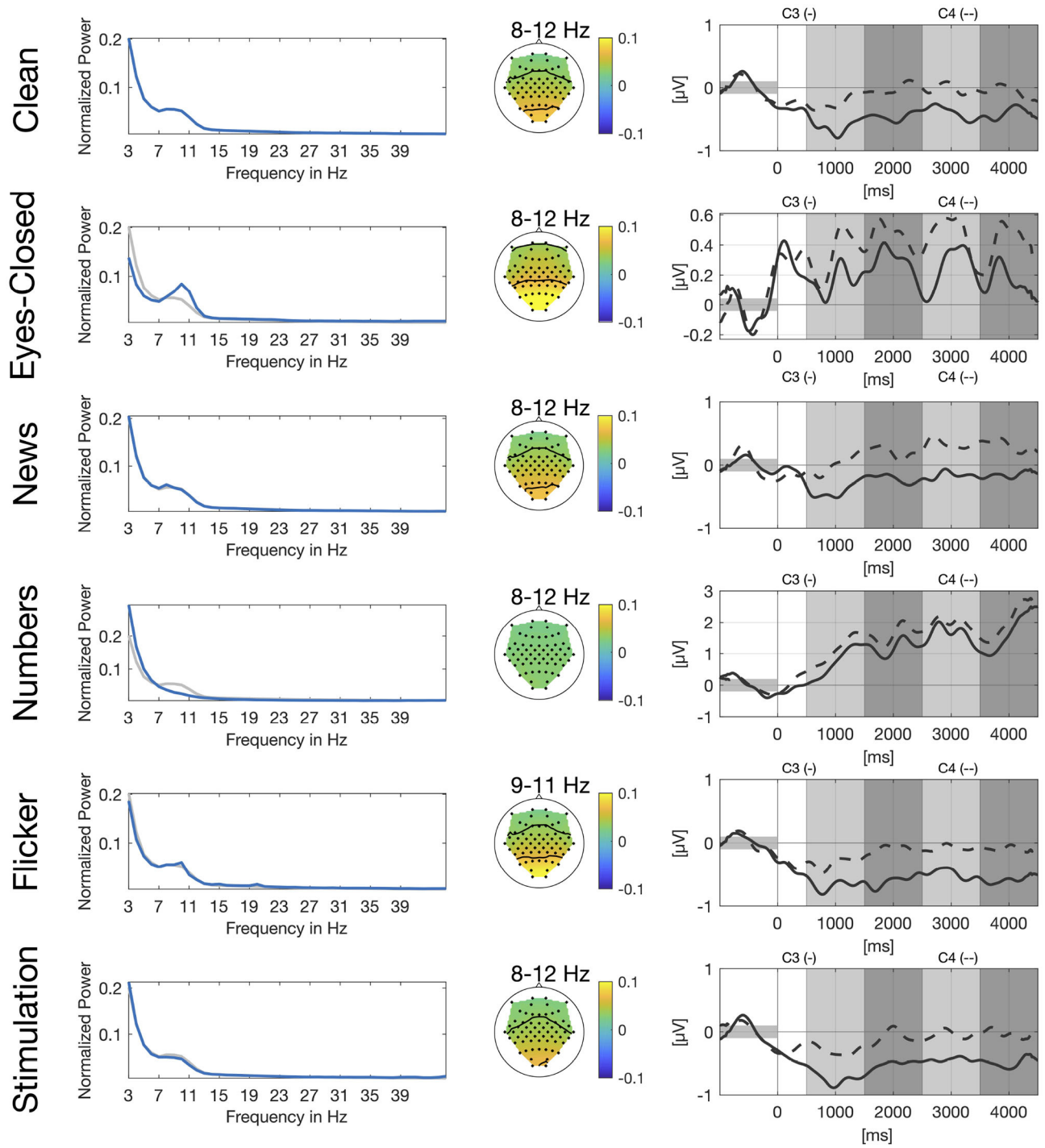


FIGURE 2 | Spectral, spatial, and temporal information for all secondary tasks. Left: Normalized power spectrum averaged across participants and channels O1 and O2, in blue for the respective secondary task and in gray for the clean condition. Center: Spatial distribution of power spectral values averaged across participants for different frequency bands. Right: Group-average envelopes in 9–13 Hz of C3 (solid) and C4 (dashed) electrodes for right hand motor imagery.

filtered in the selected frequency band with a 3rd order zero-phase Butterworth filter and cut into epochs. Six CSP filters were extracted, three per class based on the "ratio-of-median" score as described in Blankertz et al. (2007). The logarithm of the variance of the CSP-filtered signal was then used as features and fed into an RLDA classifier. Overall classification averaged across all participants reached an accuracy of 61.81%. Classification results of CSP vs. Laplacian filters are plotted in **Figure 1** (61.81

vs. 57.05%) as well as classification of CSP on *clean* condition vs. the five distraction tasks (67.08 vs. 60.76%).

In **Figure 2**, we show power spectra for all secondary tasks. For each participant, power spectra were averaged across trials and normalized channel-wise. We then extracted the power spectra for the channels O1 and O2, averaged over the two channels and again across participants. Alpha peaks clearly differ for *eyes-closed* and *numbers* compared to *clean*. For the *eyes-closed* task, we see the expected alpha peak in the range of 8–12 Hz (Berger, 1929). For the *numbers* task there is no clear alpha peak visible in the occipital channels which is in line with the expected suppression of the visual alpha rhythm during visual search. Power spectrum for the *flicker* task shows a small sharp peak between 9 and 11 Hz which is very close to the frequency of the flickering video and another even smaller peak at 20 Hz which represents the second harmonic of the flicker frequency. The *news* and *stimulation* task do not show clear differences compared to *clean*.

We also show spatial distribution for different frequency bands in the alpha range based on the peaks in the power spectrum. For *eyes-closed* and *flicker* we see a clear activation over the occipital and parietal cortex whereas there is no clear pattern visible for the *numbers* task. Again, patterns for the *news* and the *stimulation* task look very similar to the pattern of the *clean* task.

Similar to **Figure 1**, we show envelopes of channels C3 and C4 for right hand motor imagery. The modulation of the sensorimotor rhythm is still visible in all conditions as a stronger ERD in C3 compared to C4. However, the effect is obscured by the different artifacts. The disturbences are smallest in the *news, flicker* and the *stimulation* tasks due to the stationary nature of the artifacts. For the *flicker* task we still see a clear difference between both channels, whereas channels are already closer for *eyes-closed* and still even closer for the *numbers* task.

3. CONCLUSION

We recorded a motor imagery-based BCI study with 16 participants where different distraction scenarios are added as secondary tasks to systematically investigate the influence of those noise sources on the motor imagery performance. We have presented group-averages that show typical ERD/ERS effects especially during the first half of the trial over the motor cortex, typical phenomena according to the literature. We further show expected differences in power spectra for occipital channels and spatial patterns for different frequency bands in the alpha range for three of the secondary tasks. We also show classification results of a standard CSP + RLDA classification pipeline that clearly show that classification accuracy decreases in the distraction tasks. All the data[2] and the code[3] is publicly available and a more advanced analysis has been published in Brandl et al. (2016).

DATA AVAILABILITY STATEMENT

The dataset recorded for this study can be found in DepositOnce[1].

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethikkommission der Charité—Universitätsmedizin Berlin. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Design of the study by SB and BB. Recording and analysis of the study by SB. SB wrote the manuscript which was revised by BB. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by the German Ministry for Education and Research as BIFOLD—Berlin Institute for the Foundations of Learning and Data (Refs. 01IS18025A and 01IS18037A).

ACKNOWLEDGMENTS

We thank all participants for taking part in the study. We also thank J. Höhne for his support and advice on recording the study. We further thank A. von Lühmann, D. Miklody, and the reviewers for valuable comments on the manuscript and N. Koreuber for code review.

REFERENCES

Ang, K. K., Guan, C., Chua, K. S. G., Ang, B. T., Kuah, C. W. K., Wang, C., et al. (2011). A large clinical study on the ability of stroke patients to use an eeg-based motor imagery brain-computer interface. *Clin. EEG Neurosci.* 42, 253–258. doi: 10.1177/155005941104200411

Berger, H. (1929). Über das Elektroenkephalogramm des Menschen. *Archiv. Psychiatr. Nervenkrankheit.* 87, 527–570. doi: 10.1007/BF01797193

Blankertz, B., Müller, K. R., Curio, G., Vaughan, T. M., Schalk, G., Wolpaw, J. R., et al. (2004). The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials. *IEEE Trans. Biomed. Eng.* 51, 1044–1051. doi: 10.1109/TBME.2004.826692

Blankertz, B., Müller, K. R., Krusienski, D. J., Schalk, G., Wolpaw, J. R., Schlögl, A., et al. (2006). The BCI competition III: validating alternative approaches to actual BCI problems. *IEEE Trans. Neural Syst. Rehabil. Eng.* 14, 153–159. doi: 10.1109/TNSRE.2006.875642

Blankertz, B., Tangermann, M., Vidaurre, C., Fazli, S., Sannelli, C., Haufe, S., et al. (2010). The Berlin brain-computer interface: non-medical uses of BCI technology. *Front. Neurosci.* 4:198. doi: 10.3389/fnins.2010.00198

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., and Muller, K.-R. (2007). Optimizing spatial filters for robust eeg single-trial analysis. *IEEE Signal Processi. Mag.* 25, 41–56. doi: 10.1109/MSP.2008.4408441

Brandl, S., Frølich, L., Höhne, J., Müller, K.-R., and Samek, W. (2016). Brain-computer interfacing under distraction: an evaluation study. *J. Neural Eng.* 13:056012. doi: 10.1088/1741-2560/13/5/056012

Brouwer, A.-M., Snelting, A., Jaswa, M., Flascher, O., Krol, L., and Zander, T. (2017). "Physiological effects of adaptive cruise control behaviour in real

driving," in *Proceedings of the 2017 ACM Workshop on an Application-Oriented Approach to BCI Out of the Laboratory* (Limassol), 15–19. doi: 10.1145/3038439.3038441

Brouwer, A.-M., and Van Erp, J. B. (2010). A tactile P300 brain-computer interface. *Front. Neurosci*. 4:19. doi: 10.3389/fnins.2010.00019

Castermans, T., Duvinage, M., Petieau, M., Hoellinger, T., Saedeleer, C., Seetharaman, K., et al. (2011). Optimizing the performances of a P300-based brain-computer interface in ambulatory conditions. *IEEE J. Emerg. Select. Top. Circuits Syst*. 1, 566–577. doi: 10.1109/JETCAS.2011.2179421

Cho, H., Ahn, M., Ahn, S., Kwon, M., and Jun, S. C. (2017). EEG datasets for motor imagery brain-computer interface. *GigaScience* 6:gix034. doi: 10.1093/gigascience/gix034

De Vos, M., Gandras, K., and Debener, S. (2014). Towards a truly mobile auditory brain-computer interface: exploring the P300 to take away. *Int. J. Psychophysiol*. 91, 46–53. doi: 10.1016/j.ijpsycho.2013.08.010

Friedman, J. H. (1989). Regularized discriminant analysis. *J. Am. Stat. Assoc*. 84, 165–175. doi: 10.1080/01621459.1989.10478752

Han, C.-H., Kim, Y.-W., Kim, S. H., Nenadic, Z., Im, C.-H., et al. (2019). Electroencephalography-based endogenous brain-computer interface for online communication with a completely locked-in patient. *J. Neuroeng. Rehabil*. 16:18. doi: 10.1186/s12984-019-0493-0

Höhne, J., Holz, E., Staiger-Sälzer, P., Müller, K.-R., Kübler, A., and Tangermann, M. (2014). Motor imagery for severely motor-impaired patients: evidence for brain-computer interfacing as superior control solution. *PLoS ONE* 9:e104854. doi: 10.1371/journal.pone.0104854

Hwang, H.-J., Han, C.-H., Lim, J.-H., Kim, Y.-W., Choi, S.-I., An, K.-O., et al. (2017). Clinical feasibility of brain-computer interface based on steady-state visual evoked potential in patients with locked-in syndrome: case studies. *Psychophysiology* 54, 444–451. doi: 10.1111/psyp.12793

Jasper, H. (1958). The ten twenty electrode system of the international federation. *EEG Clin. Neurophysiol*. 10, 371–375.

Kaya, M., Binli, M. K., Ozbay, E., Yanar, H., and Mishchenko, Y. (2018). A large electroencephalographic motor imagery dataset for electroencephalographic brain computer interfaces. *Sci. Data* 5:180211. doi: 10.1038/sdata.2018.211

Ko, L.-W., Komarov, O., Hairston, W. D., Jung, T.-P., and Lin, C.-T. (2017). Sustained attention in real classroom settings: an EEG study. *Front. Hum. Neurosci*. 11:388. doi: 10.3389/fnhum.2017.00388

Leeb, R., Perdikis, S., Tonin, L., Biasiucci, A., Tavella, M., Creatura, M., et al. (2013). Transferring brain-computer interfaces beyond the laboratory: successful application control for motor-disabled users. *Artif. Intell. Med*. 59, 121–132. doi: 10.1016/j.artmed.2013.08.004

Lotte, F., Fujisawa, J., Touyama, H., Ito, R., Hirose, M., and Lécuyer, A. (2009). "Towards ambulatory brain-computer interfaces: a pilot study with P300 signals," in *Proceedings of the International Conference on Advances in Computer Enterntainment Technology* (Athens), 336–339. doi: 10.1145/1690388.1690452

Lugo, Z. R., Pokorny, C., Pellas, F., Noirhomme, Q., Laureys, S., Müller-Putz, G., et al. (2020). Mental imagery for brain-computer interface control and communication in non-responsive individuals. *Ann. Phys. Rehabil. Med*. 63, 21–27. doi: 10.1016/j.rehab.2019.02.005

McFarland, D. J., McCane, L. M., David, S. V., and Wolpaw, J. R. (1997). Spatial filter selection for EEG-based communication. *Electroencephalogr. Clin. Neurophysiol*. 103, 386–394. doi: 10.1016/S0013-4694(97)00022-2

Miklody, D., Uitterhoeve, W. M., Heel, D., v., Klinkenberg, K., and Blankertz, B. (2017). "Maritime cognitive workload assessment," in *Symbiotic Interaction. Symbiotic 2016. Lecture Notes in Computer Science*, Vol. 9961, eds L. Gamberini, A. Spagnolli, G. Jacucci, B. Blankertz, and J. Freeman (Cham: Springer), 102–114. doi: 10.1007/978-3-319-57753-1_9

Morgan, S., Hansen, J., and Hillyard, S. (1996). Selective attention to stimulus location modulates the steady-state visual evoked potential. *Proc. Natl. Acad. Sci. U.S.A*. 93, 4770–4774. doi: 10.1073/pnas.93.10.4770

Neuper, C., Müller, G., Kübler, A., Birbaumer, N., and Pfurtscheller, G. (2003). Clinical application of an EEG-based brain-computer interface: a case study in a patient with severe motor impairment. *Clin. Neurophysiol*. 114, 399–409. doi: 10.1016/S1388-2457(02)00387-5

Pfurtscheller, G. (1992). Event-related synchronization (ERS): an electrophysiological correlate of cortical areas at rest. *Electroencephalogr. Clin. Neurophysiol*. 83, 62–69. doi: 10.1016/0013-4694(92)90133-3

Ramoser, H., Muller-Gerking, J., and Pfurtscheller, G. (2000). Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans. Rehabil. Eng*. 8, 441–446. doi: 10.1109/86.895946

Sajda, P., Gerson, A., Müller, K.-R., and Blankertz, B. (2003). A data analysis competition to evaluate machine learning algorithms for use in brain-computer interfaces. *Rehabilitation* 11, 184–185. doi: 10.1109/TNSRE.2003.814453

Shin, J., von Lühmann, A., Blankertz, B., Kim, D.-W., Jeong, J., Hwang, H.-J., et al. (2016). Open access dataset for EEG + NIRS single-trial classification. *IEEE Trans. Neural Syst. Rehabil. Eng*. 25, 1735–1745. doi: 10.1109/TNSRE.2016.2628057

Tangermann, M., Müller, K. R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., et al. (2012). Review of the BCI competition IV. *Front. Neurosci*. 6:55. doi: 10.3389/fnins.2012.00055

Tobimatsu, S., Zhang, Y. M., and Kato, M. (1999). Steady-state vibration somatosensory evoked potentials: physiological characteristics and tuning function. *Clin. Neurophysiol*. 110, 1953–1958. doi: 10.1016/S1388-2457(99)00146-7

Vidaurre, C., Kawanabe, M., von Bünau, P., Blankertz, B., and Müller, K.-R. (2010). Toward unsupervised adaptation of LDA for brain-computer interfaces. *IEEE Trans. Biomed. Eng*. 58, 587–597. doi: 10.1109/TBME.2010.2093133

von Lühmann, A., Boukouvalas, Z., Müller, K.-R., and Adalı, T. (2019). A new blind source separation framework for signal analysis and artifact rejection in functional near-infrared spectroscopy. *Neuroimage* 200, 72–88. doi: 10.1016/j.neuroimage.2019.06.021

Wriessnegger, S. C., Krumpl, G., Pinegger, A., and Mueller-Putz, G. (2017). "Mobile BCI technology: NeuroIS goes out of the lab, into the field," in *Information Systems and Neuroscience*, eds F. Davis, R. Riedl, J. vom Brocke, P. M. Léger, and A. Randolph (Cham: Springer), 59–65. doi: 10.1007/978-3-319-41402-7_8

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 22 October 2020
doi: 10.3389/fnins.2020.579469

A Cross-Session Dataset for Collaborative Brain-Computer Interfaces Based on Rapid Serial Visual Presentation

*Li Zheng[1,2], Sen Sun[3], Hongze Zhao[1], Weihua Pei[1,2], Hongda Chen[1], Xiaorong Gao[4], Lijian Zhang[5] and Yijun Wang[1,2]**

[1] State Key Laboratory on Integrated Optoelectronics, Institute of Semiconductors, Chinese Academy of Sciences, Beijing, China, [2] School of Future Technology, University of Chinese Academy of Sciences, Beijing, China, [3] Department of Control Engineering, School of Information Science and Engineering, East China University of Science and Technology, Shanghai, China, [4] Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing, China, [5] Beijing Machine and Equipment Institute, Beijing, China

OPEN ACCESS

Edited by:
Ana Matran-Fernandez, University of Essex, United Kingdom

Reviewed by:
Alan F. Smeaton, Dublin City University, Ireland
Hubert Cecotti, California State University, Fresno, United States

***Correspondence:**
Yijun Wang
wangyj@semi.ac.cn

Specialty section:
This article was submitted to Neural Technology, a section of the journal Frontiers in Neuroscience

Received: *02 July 2020*
Accepted: *22 September 2020*
Published: *22 October 2020*

Citation:
Zheng L, Sun S, Zhao H, Pei W, Chen H, Gao X, Zhang L and Wang Y (2020) A Cross-Session Dataset for Collaborative Brain-Computer Interfaces Based on Rapid Serial Visual Presentation. Front. Neurosci. 14:579469. doi: 10.3389/fnins.2020.579469

Brain-computer interfaces (BCIs) based on rapid serial visual presentation (RSVP) have been widely used to categorize target and non-target images. However, it is still a challenge to detect single-trial event related potentials (ERPs) from electroencephalography (EEG) signals. Besides, the variability of EEG signal over time may cause difficulties of calibration in long-term system use. Recently, collaborative BCIs have been proposed to improve the overall BCI performance by fusing brain activities acquired from multiple subjects. For both individual and collaborative BCIs, feature extraction and classification algorithms that can be transferred across sessions can significantly facilitate system calibration. Although open datasets are highly efficient for developing algorithms, currently there is still a lack of datasets for a collaborative RSVP-based BCI. This paper presents a cross-session EEG dataset of a collaborative RSVP-based BCI system from 14 subjects, who were divided into seven groups. In collaborative BCI experiments, two subjects did the same target image detection tasks synchronously. All subjects participated in the same experiment twice with an average interval of ∼23 days. The results in data evaluation indicate that adequate signal processing algorithms can greatly enhance the cross-session BCI performance in both individual and collaborative conditions. Besides, compared with individual BCIs, the collaborative methods that fuse information from multiple subjects obtain significantly improved BCI performance. This dataset can be used for developing more efficient algorithms to enhance performance and practicality of a collaborative RSVP-based BCI system.

Keywords: brain-computer interfaces (BCI), rapid serial visual presentation (RSVP), collaborative BCI, cross-session transfer, event related potentials (ERP), electroencephalogram (EEG)

INTRODUCTION

Brain-computer interfaces (BCIs) establish a communication channel between human brain and the external world (Wolpaw et al., 2002; Gao et al., 2014). As one of the well-known BCI paradigms, rapid serial visual presentation (RSVP)-based BCIs have been usually used for target image detection. Although computer vision (CV) has become a major method to deal with the image recognition problem recently, it consumes a large amount of resource (image source, training time, computing power, etc.) to get a good performance, and is still lack of generalization ability. By contrast, human vision (HV) can achieve general purposes of object recognition. HV can cope with more difficult tasks and detect targets with different characteristics (e.g., scale, lighting, background, etc.; Mathan et al., 2008; Sajda et al., 2010; Pohlmeyer et al., 2011). Human visual system can recognize objects with just a glance (Oliva, 2005) and detect targets in under 150 ms (Thorpe et al., 1996). However, manual image analysis is slow because of the motor response time, and the variability of response time makes it difficult to locate the target images in RSVP tasks (Gerson et al., 2005; Mathan et al., 2008; Sajda et al., 2010). Therefore, RSVP-based BCIs, which have stronger generalization ability and are faster than behavioral response, have become a useful method to detect targets by using the human brain activities. By presenting multiple images sequentially in a high presentation rate (e.g., 10 images per second), the RSVP-based BCI can enhance the target detection performance of HV (Lees et al., 2018).

In earlier times, RSVP was often used to do behavioral research focusing on attentional blink (AB; Broadbent and Broadbent, 1987; Chun and Potter, 1995; Jolicoeur, 1998) and manual target detection (Lawrence, 1971; Broadbent and Broadbent, 1987). With the rapid development of computer technology and electroencephalography (EEG)-based BCIs, RSVP was introduced to design BCI systems for target detection. The RSVP-based BCI is realized by single-trial event related potential (ERP) detection. ERPs typically contain multiple components with different temporal and spatial characters. In an RSVP-based BCI system, the P300 component, which occurs approximately 300 ms after the target stimulation, is the major ERP component used for target detection (Picton, 1992; Chun and Potter, 1995). Since the system performance of RSVP-based BCIs can be influenced by many factors such as presentation rate (Sajda et al., 2003; Acqualagna et al., 2010; Lees et al., 2019), target probability (Cecotti et al., 2011), stimulus onset asynchrony and stimulus repetition (Cecotti et al., 2014a), image size (Rousselet et al., 2004; Serre et al., 2007), type of targets (Lees et al., 2019), saccadic eye movements (Bigdely-Shamlo et al., 2008), attention blink (Broadbent and Broadbent, 1987; Chun and Potter, 1995; Jolicoeur, 1998), and other subjective or objective factors (Jolicoeur, 1998; Acqualagna et al., 2010; Touryan et al., 2011), the experimental paradigm should be carefully designed.

Besides the design of system paradigm, the main challenge in RSVP-based BCIs is single-trial ERP detection. In the RSVP-based BCI system, multi-channel EEG recording leads to a high dimensionality of features, and the small number of trials is always not sufficient for solving the classification problem toward accurate ERP detection (Huang et al., 2011). To deal with the problem of single-trial ERP detection, suitable signal processing and classification algorithms are required to extract discriminative information from single-trial data and improve the performance in classifying target and non-target trials. Various algorithms have been proposed and developed for the RSVP-based BCIs (Lees et al., 2018; Lotte et al., 2018). Major feature extraction algorithms include xDAWN (Rivet et al., 2009), signal-to-noise ratio (SNR) maximizer for ERP (SIM; Wu and Gao, 2011), common spatial pattern (CSP; Ramoser et al., 2000), independent component analysis (ICA; Makeig et al., 1996), and etc. Typical classification algorithms include spatially weighted fisher's linear discriminant (FLD)-principal component analysis [PCA; spatially weighted FLD-PCA (SWFP); Alpert et al., 2013], support vectors machine (SVM; Burges, 1998), linear discriminate analysis (LDA; Blankertz et al., 2011), hierarchical discriminant component analysis (HDCA; Sajda et al., 2010), convolutional neural network (CNN; Cecotti and Graser, 2010; Cecotti et al., 2014b), and etc. Since real targets can only appear once in the RSVP paradigm, averaging across multiple trials is not practical in the RSVP-based BCIs. By combining brain activities of multiple subjects, collaborative BCIs can improve the performance of single-trial ERP detection (Wang and Jung, 2011). A series of studies have demonstrated collaborative BCIs for target detection and decision making (Wang et al., 2011; Yuan et al., 2012; Matran-Fernandez et al., 2013; Cecotti and Rivet, 2014; Poli et al., 2014; Touyama, 2014; Valeriani et al., 2015, 2016, 2017; Bhattacharyya et al., 2019). For both individual and collaborative RSVP-based BCIs, system calibration remains another big challenge in practical applications. It has been claimed that high variability of EEG makes it difficult to transfer models across different sessions (Krauledat et al., 2008). Besides, the training session in system calibration is time-consuming and the system performance may probably decrease over time (Bigdely-Shamlo et al., 2008; Huang et al., 2011; Zhao et al., 2019). Therefore, it is of great significance to develop efficient algorithms to solve the cross-session classification problem in the RSVP-based BCIs.

Recently, open BCI datasets have pushed forward the development of data processing algorithms. However, there are very few freely available datasets for the RSVP-based BCIs (Acqualagna and Blankertz, 2013; Matran-Fernandez and Poli, 2017). To our knowledge, a benchmark dataset for collaborative RSVP-based BCIs is still missing. Besides, the existing datasets only provide data recorded from a single session, which is not suitable for studying the problem of cross-session transfer. This paper therefore presents a cross-session dataset for collaborative RSVP-based BCIs. The dataset has the following characteristics: (1) EEG data from two subjects were recorded simultaneously with a collaborative BCI where two subjects performed the same target detection tasks synchronously, (2) two separate sessions were recorded for each of seven groups (14 subjects) on two different days with an average interval of ∼23 days, and (3) whole-head 62-channel EEG data were recorded and the raw data were provided without further processing. Note that, all event triggers for target and non-target images were synchronously marked in the EEG data. Therefore, the data epochs extracted

from both subjects could be precisely synchronized. During the experiments, subjects were asked to find target images with human in street images sequences presented at 10 Hz (10 images per second). The experiment included three blocks, and each block contained 14 trials. Each trial had 100 images, including 4 target images. In total, the dataset contains 84 blocks (1,176 trials) of data recorded from 14 subjects. The dataset can be especially useful for studying cross-session ERP detection algorithms for both individual and collaborative RSVP-based BCI systems.

The rest of this paper is organized as follows. Section "Methods" explains the experimental paradigm, data acquisition, the algorithms in data analysis, and the criterion in performance evaluation. Section "Data Record" describes the data record and other relevant information. Section "Data Evaluation" presents results of BCI performance in data evaluation. Section "Conclusion and Discussions" concludes and discusses future works.

METHODS

Participants

Fourteen healthy subjects (10 females, mean age: 24.9 ± 1.5 years, all right-handed) with normal or corrected-to-normal vision participated in the experiments. The subjects were divided into seven groups with two subjects in each group. For each group, the experiments contained two sessions recorded on different days. For all groups, the average time interval between two sessions was ∼23 days. All subjects were asked to read and sign an informed consent form before the experiment. This study was approved by the Ethics Committee of Tsinghua University.

Collaborative System

Figure 1 illustrates the diagram of the online collaborative BCI system. The system consists of four major components: Stimulation module, Operation module, Data Acquisition module, and Command and Data Analysis module. The system performs the following steps: (1) The Command and Data Analysis module waits for keypress information from the Stimulation modules to start a trial; (2) The Command and Data Analysis module sends synchronous commands to the Operation and Stimulation modules; (3) The Stimulation modules present the RSVP stimuli to the subjects and (4) send event triggers to the Data Acquisition modules; (5) The Operation modules send control commands to the Data Acquisition modules and (6) record EEG data from the subjects; (7) The Operation modules receive EEG data from the Data Acquisition modules and (8) transfer to the Command and Data Analysis module; (9) The Command and Data Analysis module analyzes data and outputs online collaborative decisions. Data packages and commands are sent using transmission control protocol/internet protocol (TCP/IP) and triggers are sent using parallel ports. In the collaborative experiment, two subjects watched the same RSVP stimuli synchronously, and EEG data from them were fused to improve the overall detection performance. The same stimulations were presented to the two subjects using two separate computers. To synchronize EEG data from the two subjects, event triggers from the two stimulation computers were sent separately. The Command and Data Analysis module sent messages to synchronize the other modules. Therefore, although the Stimulation, Operation, and Data Acquisition modules were separated for each subject, the Command and Data analysis module fused the EEG data from two subjects and performed collaborative target detection in real time.

Collaborative Experiment Design

The stimulation pattern of the RSVP paradigm is shown in **Figure 2**. The stimulation is presented by a 24.5-inch liquid crystal display (LCD) monitor with a resolution of 1,920 × 1,080 and a vertical refresh rate of 60 Hz. The images were downloaded

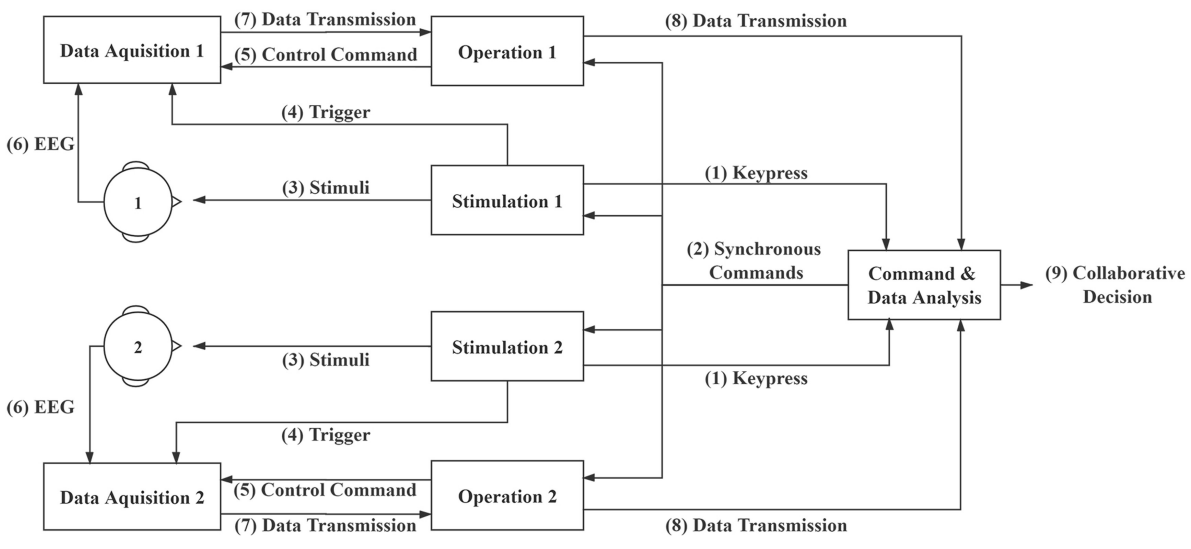


FIGURE 1 | System diagram of the collaborative brain-computer interfaces (BCI) system.

from the internet. The stimulation was generated using the Psychophysics Toolbox Ver. 3 (PTB-3; Brainard, 1997). Street scene images were presented at 10 Hz (10 images per second) in the center of the screen within a 1,200 × 800-pixel square. The images containing human were regarded as target images.

The procedures of the collaborative experiments are depicted as follows. Subjects were asked to sit comfortably approximately 70 cm in front of the screen. When the subjects were ready, both of them were supposed to press keys to start one trial. The stimulation would not begin until both subjects pressed keys. If one subject pressed the key first, he or she had to wait for the second subject's keypress to start the trial at the same time. After receiving two keypresses, the command module sent commands to the stimulation modules to start the same image sequence presentation synchronously to two subjects. As shown in **Figure 2**, a cross symbol appeared at the center of the screen for 500 ms to make subjects fix their sights, then the RSVP stimulation began. Each trial contained 100 images (10 s at the rate of 10 Hz), including four target images. The images shown in the first and last 1 s in one trial were all non-target images to avoid the target from appearing during the onset or offset of steady-state visual evoked potentials (SSVEP) evoked by RSVP. The interval of two target images was at least 500 ms to reduce the influence of the attention blink (Broadbent and Broadbent, 1987; Chun and Potter, 1995; Lees et al., 2018). Subjects were asked to press keys immediately after they detected a target. The keypress task was used to make subjects concentrate on target detection. Since there was a time delay between the target image and keypress, the keypress within 500ms after a target image was considered a correct response to the target image during the experiments. In the experiments, subjects needed to find four targets from 100 images and made four keystrokes. If the subjects missed some targets, the system would show the missed targets at the end of the trial. For the same group of subjects, the experiments included two sessions on different days, where the stimulation paradigms were totally same. The RSVP stimulation was presented in blocks. Each session consisted of three blocks and each block contained 14 trials (1,400 images, including 56 targets). Subjects were allowed to take a short rest after each block. During the experiment, the first block was used for training, while the second and the third blocks were used for testing. In the testing blocks, online classification results were provided by the Command and Data Analysis module. The online visual feedback was a 3 × 3 image matrix including nine images with the highest scores among the 100 images in each trial.

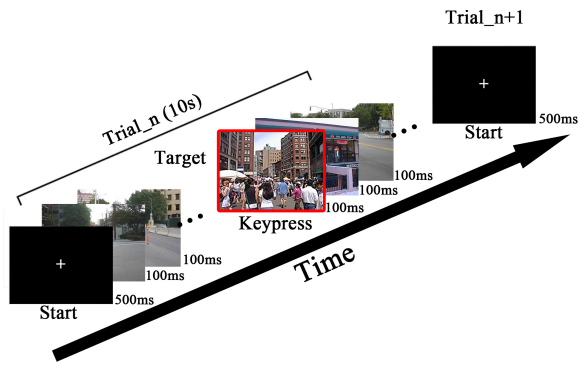


FIGURE 2 | The overview of rapid serial visual presentation (RSVP) stimulation. The images are presented at 10 Hz. Subjects were asked to press the key immediately when finding a target. The sample of target is highlighted with a red frame.

Data Acquisition

The EEG data from two subjects were simultaneously recorded by two Neuroscan Synamps2 systems. 64-electrode EEG caps based on the 10–20 system were used to record 62-channel EEG data (M1 and M2 were not used) from two subjects. The reference electrode was at the vertex. The impedances of the electrodes were kept under 10 kΩ. The sample rate was 1,000 Hz. A notch filter at 50 Hz was used to remove the common power-line noise. The pass-band of the amplifier was set to 0.15–200 Hz. All the event triggers were transmitted and marked on the EEG data by parallel ports. Two stimulation computers sent triggers separately to the two EEG systems. The dataset provides raw data from the experiments without any processing.

Data Preprocessing

To validate the quality of the data through performance evaluation, data preprocessing was performed as follows. The EEG data were first down-sampled to 250 Hz. After that, epochs corresponding to all images were extracted according to the event triggers. Each epoch began at 0.2 second before the event trigger, and ended at 1 second after the event trigger. The epochs were band-pass filtered within 2–30 Hz. For the analysis of EEG characteristics, the EEG data were re-referenced to the average of all electrodes [i.e., common average reference (CAR)], and the ERP waveforms were plotted using data at Cz. For performance evaluation, the time window 0–500 ms after the event trigger of each epoch was extracted for feature extraction and classification.

Data Analysis

Individual Data Analysis

In this paper, several existing algorithms were utilized for feature extraction and classification. The HDCA algorithm, which can extract both spatial and temporal features, has been widely used in the RSVP-based BCIs (Lees et al., 2018; Zhao et al., 2019; Sajda et al., 2010). In our previous study, the combination of SIM and HDCA was employed to deal with the cross-session transfer problem (Zhao et al., 2019). SIM can extract the EEG components that maximize the SNR of ERPs (Wu and Gao, 2011). In this paper, several other feature extraction algorithms including CSP, task-related component analysis (TRCA), and PCA whitening were employed for comparison. CSP can build a spatial filter to extract features from two classes toward the best discrimination (Ramoser et al., 2000). TRCA is a method to extract task-related components by maximizing the reproducibility of repetitive tasks (Nakanishi et al., 2018). PCA whitening is usually used before ICA to reduce the complexity of the classification problem (Hyvärinen and Oja, 2000). To estimate performance for each subject, the first block of data was used for training and the other two blocks were used for testing.

Collaborative Data Analysis

The diagrams of collaborative data analysis are depicted in **Figure 3**. For the collaborative experiments, the EEG data of two subjects were fused by three methods: ERP averaging, feature concatenating, and voting (Wang and Jung, 2011). ERP averaging and feature concatenating are centralized methods, which fuse the data before further feature extraction and classification algorithms. Voting is a distributed method, which analyzes data of each subject first and then fuses the scores generated by the individual classifiers. In the ERP averaging method, the synchronous data epochs of two subjects were averaged. In the feature concatenating method, data epochs of two subjects were concatenated for further analysis. In the voting method, the weighted sum of the output scores of the classifiers of two subjects were used for classification, and the weights were the performance [i.e., area under curve (AUC)] of each subject from the training procedure. During the experiments, the online feedback, which consisted of nine images with the highest output scores, was calculated using the voting method.

Cross-Session Data Analysis

For the cross-session data analysis, the algorithms used for evaluation were the same as the separate experiments. However, the number of components extracted by the feature extraction algorithms (e.g., spatial filtering methods such as CSP, TRCA, and SIM) was optimized separately for each algorithm. The number of components can influence the cross-session performance because of the cross-session variability of EEG data. To estimate the cross-session performance, the first block of data on Day 1 was used for training and the second and third blocks on Day 2 were used for testing. The validation strategy was the same for individual and collaborative data analysis.

Metric

This paper used the area under receiver operating characteristic (ROC) to evaluate the BCI performance. This metric is suitable for the RSVP paradigm where the class distribution is unbalanced (Lees et al., 2018). AUC can reflect the relationship between true positive rate (TPR) and false positive rate (FPR). In the RSVP-based BCI system, higher AUC indicates better performance.

DATA RECORD

EEG Data

The dataset is freely available at https://doi.org/10.6084/m9.figshare.12824771.v1. The dataset is about 6.58 GB including collaborative and cross-session data from 14 subjects. All data are saved as MATLAB MAT files. The sample rate is 1,000 Hz and all data are raw data without any processing. Each file is named as "Group index + Session index" (i.e., G1D1.mat, G1D2.mat, . . ., G7D2.mat). "G*n*" is the *n*th group (totally seven). "D1" and "D2" indicate the first and second sessions respectively. Each file contains two cells named "Sa" and "Sb" indicating two subjects in the group. Each 1 × 3 cell array ("Sa" and "Sb") contains three blocks of data recorded in one session. Each element in the cell array corresponds to one block of data. Each element is a matrix with a dimension of [63, *N*], which indicates 62-channel EEG data and a trigger channel with a length of *N*. *N* of each matrix is different because of the different experiment duration, but *N* of a group of subjects in the same block is the same. For the trigger channel, the onset of target image is defined as "1" and the onset of non-target image is defined as "2." Since each element corresponds to one block, each matrix contains data of 14 trials (1,400 image events, including 56 targets). Details of data information are also summarized in a "Readme.txt" file.

Supplementary Information

Three supplementary files are provided including subject information and channel location, and the image set. Subject information is saved in a "sub_info.txt" file, which includes the gender, age, handedness, group, and the interval between the two sessions. Channel locations are saved in a "62-channels.loc" file, in which the information for each channel consists of four columns: channel index, degree, radius, and label. The origin is at Cz (i.e., the radius is 0). The image set used for RSVP stimulation is also included in the supplementary files and

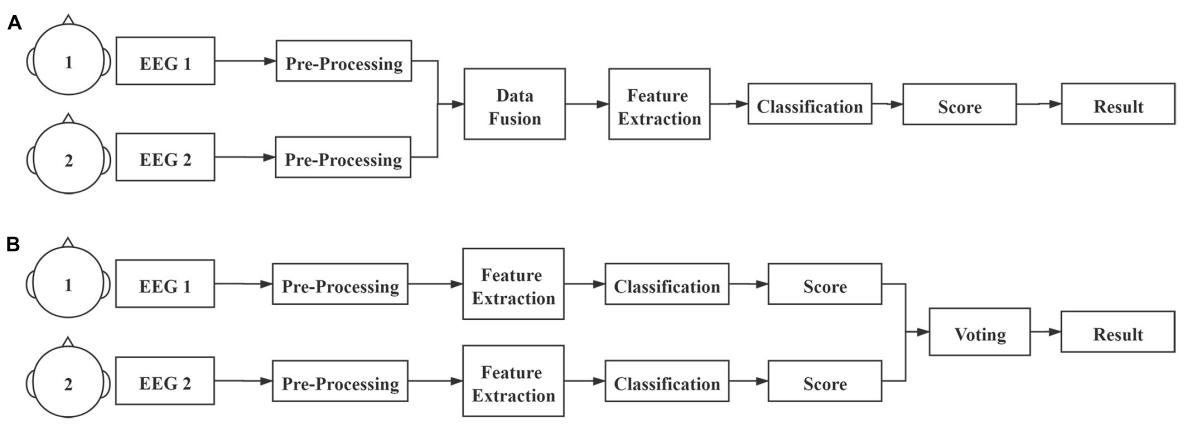


FIGURE 3 | (A) Centralized and **(B)** distributed diagrams of collaborative data analysis.

saved in a "Image.zip" file. Target and non-target images are saved in two folders.

DATA EVALUATION

Individual BCI Performance

Within-Session Individual Performance

Figure 4 shows the EEG characteristics on Day 1 (D1) and Day 2 (D2) related to targets and non-targets using the average data of all subjects. **Figure 4A** shows the time course of scalp map series of average ERP amplitudes. It is clearly shown that the P300 component peaked around 400 ms and was mainly distributed at the central-parietal areas. **Figure 4B** shows the average ERP waveform at Cz. The waveform shows large N2 and P3 components after the target onset, which are obviously higher than other ERP components. **Figure 4C** shows the spectrum of ERPs at Cz calculated by fast Fourier transform (FFT). The EEG power mainly focuses at a low frequency range under 10 Hz with a peak around 4 Hz. It should be noted that there are frequency peaks at 10 Hz and its harmonic frequencies, which means there are SSVEP components in ERPs. **Figures 4D–F** show the EEG characteristics related to the non-target images. **Figure 4D** shows the average topographic map series, which indicate significant distribution of SSVEPs mainly focused at the occipital area. The average EEG waveform at Cz in **Figure 4E** indicates strong SSVEP components evoked by the 10 Hz RSVP stimulation. **Figure 4F** shows the spectrum of the average EEG waveform at Cz with peaks at 10 Hz and its harmonic frequencies. In summary, EEG signals are different when subjects watch target and non-target images. During the RSVP task, SSVEP components are dominant when there are no targets, while the ERP components (i.e., N2 and P3) are evoked when detecting a target. For target images, the amplitude of ERPs are significantly higher than SSVEPs (**Figure 4C**). The amplitudes of SSVEPs for target and non-target images are close (**Figures 4C,F**).

The within-session BCI performance of individual classification is illustrated in **Figure 5**. For each subject, the first block of data is used for training and the other two blocks are used for testing. Four feature extraction methods are combined with HDCA for comparison: (1) CSP, (2) TRCA, (3)

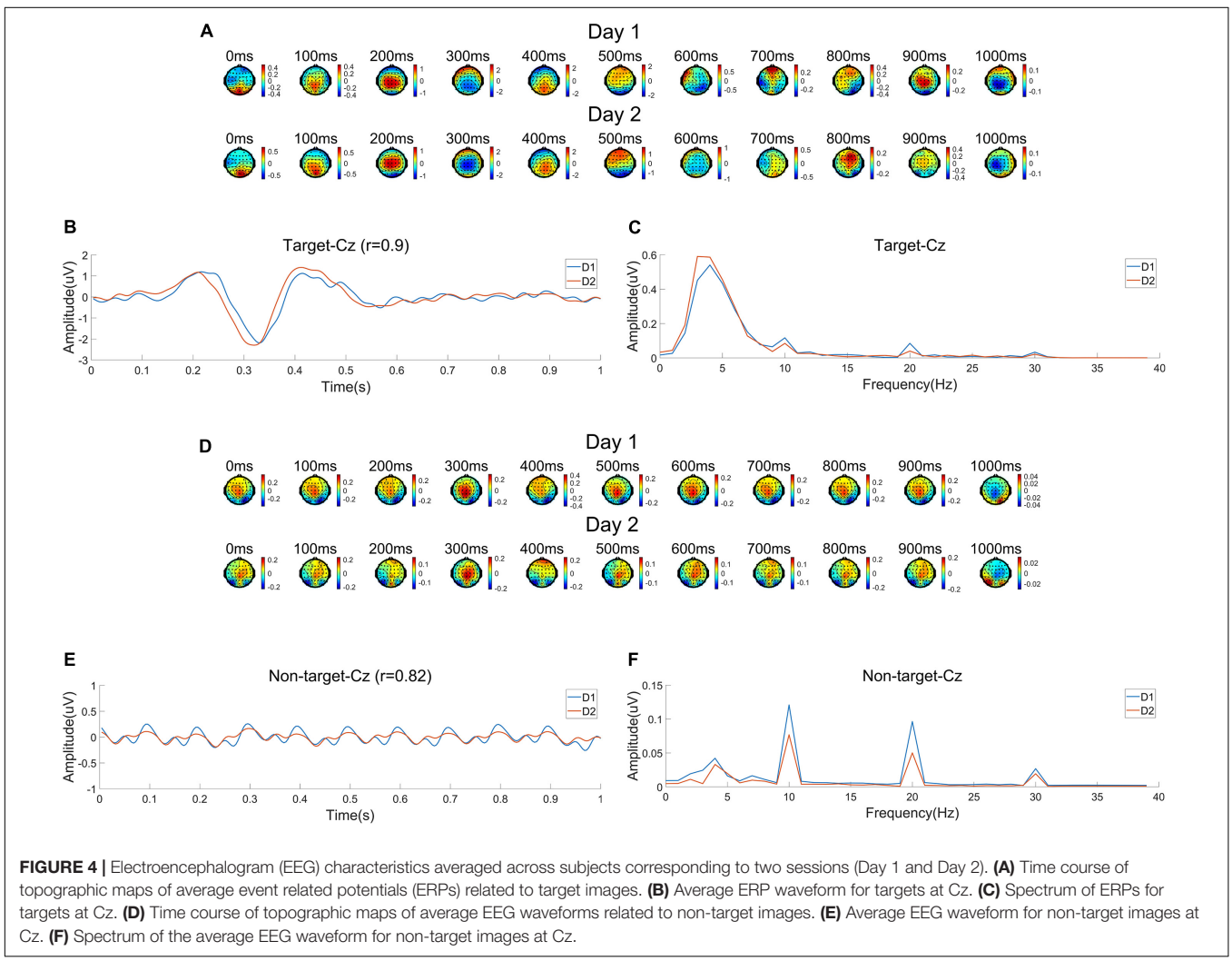


FIGURE 4 | Electroencephalogram (EEG) characteristics averaged across subjects corresponding to two sessions (Day 1 and Day 2). **(A)** Time course of topographic maps of average event related potentials (ERPs) related to target images. **(B)** Average ERP waveform for targets at Cz. **(C)** Spectrum of ERPs for targets at Cz. **(D)** Time course of topographic maps of average EEG waveforms related to non-target images. **(E)** Average EEG waveform for non-target images at Cz. **(F)** Spectrum of the average EEG waveform for non-target images at Cz.

PCA, and (4) SIM. Because the number of components (i.e., the number of spatial filters) can influence the performance, all the component numbers are calculated, and the number of components with the maximum AUC values is shown in **Figure 5**. For each feature extraction method, the numbers of used components in the within-session conditions (D1–D1, D2–D2) and the cross-session condition (D1–D2) are different (CSP: 50, 50, 13; TRCA: 7, 3, 3; PCA: 24, 36, 14; SIM: 24, 36, 21). For the within-session classification, the combination methods cannot outperform the standard HDCA algorithm. Feature extraction algorithms can help to enhance the average performance (e.g., max ΔAUC = 0.018 when using the SIM + HDCA method on Day 1), but the improvement is not substantial.

Cross-Session Individual Performance

The cross-session variances of EEG are illustrated in **Figure 4**. In **Figure 4A**, ERP scalp map series for the two sessions (Day 1 and Day 2) show similar spatial and temporal trends, but also indicate different amplitudes and spatial distributions. In **Figure 4B**, differences in amplitudes and latencies of ERP components (i.e., N2 and P3) can be observed. The correlation of two average ERP waveforms in the two sessions still obtains a high correlation coefficient of 0.90. The spectral distributions shown in **Figure 4C** are consistent for the two sessions. The cross-session differences of SSVEPs related to the non-target images can be observed in **Figures 4D–F**. There are cross-session differences in terms of spatial distributions, amplitudes and latencies. The correlation coefficient of waveforms is 0.82. Although peaks of the spectra in both sessions are at 10 Hz and its harmonic frequencies, the peak amplitudes in the two sessions seem different in **Figure 4F**. The consistency of ERP characteristics across sessions suggests it is possible to transfer information from a previous session to facilitate system calibration.

The cross-session BCI performance of individual experiments is illustrated in **Figure 5**. Compared with the within-session performance, the cross-session performance decreases sharply due to the non-stationarity of EEG over time. For example, the AUC for HDCA decreases from 0.90 (Day 1 and Day 2) to 0.67. As described above, feature extraction algorithms do not improve the within-session performance, but the PCA and SIM algorithms significantly improve the cross-session classification performance. As shown in **Figure 5**, AUC values of the SIM + HDCA (0.86 $\pm$ 0.06) and PCA + HDCA (0.86 $\pm$ 0.06) methods are significantly better than HDCA (0.67 $\pm$ 0.11), CSP + HDCA (0.69 $\pm$ 0.11) and TRCA + HDCA (0.71 $\pm$ 0.13; $p < 0.001$). There is no significant difference between SIM + HDCA and PCA + HDCA ($p > 0.05$). **Figure 6** shows an example of the relationship between AUC and the number of components when using the combination of SIM and HDCA algorithms. For the within-session condition, AUC increases as the number of components in SIM increases until AUC saturates when the number of components increases to about 30. However, for the cross-session condition, AUC first increases from 0.74 with 1 component, reaches a peak value of 0.87 with 21 components, and then decreases to 0.67 with 62 components. This finding indicates that not all spatial filters are suitable for cross-session transfer. **Figure 7** shows averaged waveforms of the 1st–20th components from one subject extracted by SIM in the cross-session condition. It is clear that the first several components show strong ERP components. When the spatial filters obtained from Day 1 are directly applied to Day 2, the first several components have high cross-session correlation coefficients. For example, for the 1st–5th components, the correlation coefficients are 0.98, 0.98, 0.93, 0.93, and 0.73 respectively. However, the correlation coefficients decrease at the latter components (e.g., the correlation coefficients are less than 0.5 for the 9th–20th components). Therefore, the first several components that show stable ERP characters in both sessions contribute most to the cross-session classification. The involvement of latter components that exhibit large difference between two sessions cannot improve the classification performance. On the contrary, the increase of feature dimension might increase

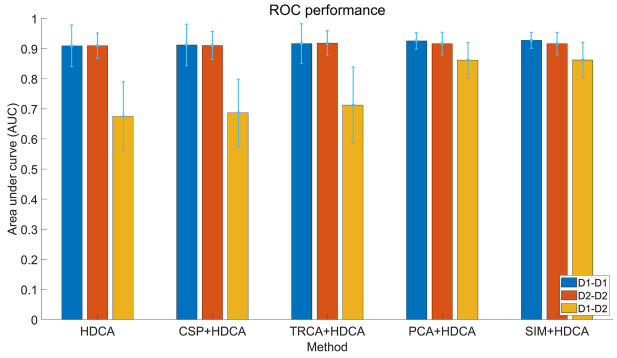


FIGURE 5 | Receiver operating characteristic (ROC) performance of different feature extraction and classification algorithms in cross-session analysis with individual classification. The color bars of each method indicate performance of within-session (Day 1 and Day 2) and cross-session classification. The error bars indicate standard deviations.

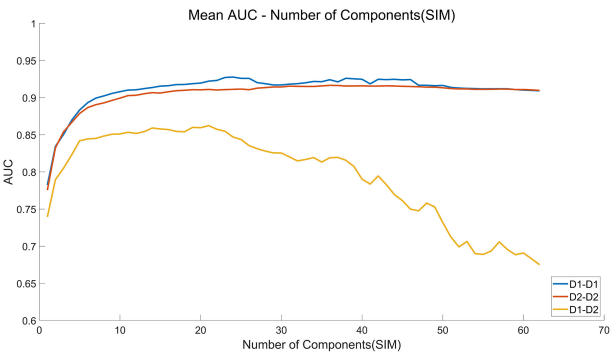


FIGURE 6 | Number of components-area under curve (AUC) curve of the SIM + hierarchical discriminant component analysis (HDCA) algorithm in cross-session individual data analysis. The *x* axis indicates the number of components selected from the spatial filters calculated by the SIM algorithm.

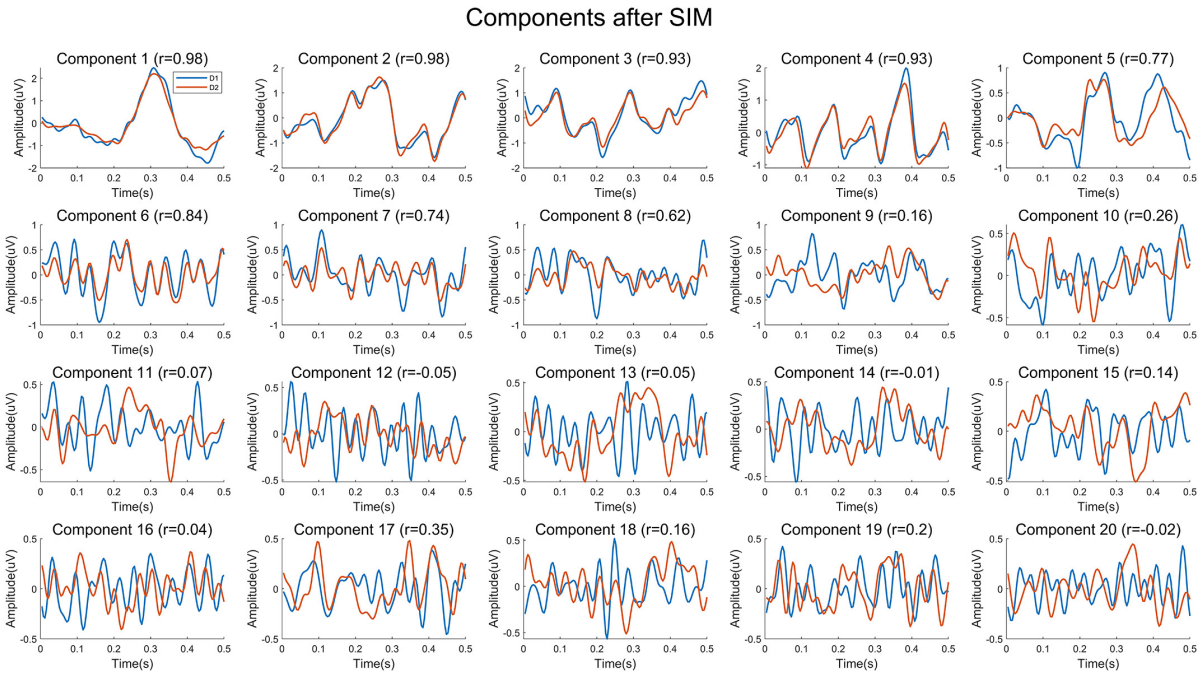


FIGURE 7 | Average waveforms of components extracted by the SIM algorithm of one subject. The waveforms indicate the 1st–20th components of Day 1 and Day 2 extracted by the spatial filters obtained from data on Day 1.

the risk of overfitting and thereby deteriorates the cross-session performance.

Collaborative BCI Performance

Within-Session Collaborative Performance

Example data recorded on Day 1 from subject 1 (Sub1) and subject 2 (Sub2) are used to analyze the EEG features for different subjects. The preprocessing procedures were the same as those in **Figure 4**. **Figure 8** illustrates the common and different characters of EEG signals for Sub1 and Sub2. Specifically, **Figures 8A–C** show the ERP characteristics related to target images. As shown in **Figure 8A**, the time courses of topographic map series for both subjects show generally similar patterns. During about 200–400 ms after the target onset, N2 and P3 components are dominant over the central-parietal areas. There is clear difference in amplitudes and latencies for ERP waveforms at Cz (**Figure 8B**), which leads to a correlation coefficient of 0.83. **Figure 8C** shows the spectral distributions of ERPs at Cz. EEG powers for the two subjects are mainly under 10 Hz with slight difference, and the amplitudes of SSVEP components are very different. **Figures 8D–F** show the EEG characteristics related to non-target images. For SSVEP components evoked by non-target images, individual difference can be observed regarding to scalp topographic maps, amplitudes and latencies, as well as spectral distributions. As shown in **Figure 8E**, the correlation coefficient of EEG waveforms is 0.45. The low correlation can be attributed to the amplitude and latency difference of the fundamental and harmonic SSVEP components shown in **Figures 8E,F**. These results suggest that, as expected, the collaborative classification will improve the individual classification by fusing useful information from multiple subjects. However, the individual difference should be carefully considered in designing the data fusion method.

The results of collaborative BCI performance are illustrated in **Figure 9**. For each group, the first block of data is used for training and the other two blocks are used for testing. The classification algorithm is SIM + HDCA (m = 30, according to the individual results in **Figure 6**). The feature fusion methods include ERP averaging, feature concatenating, and voting. As shown in **Figure 9**, all feature fusion algorithms can significantly improve the average individual performance (Single subject: 0.91 ± 0.03, ERP averaging: 0.94 ± 0.04, Feature concatenating: 0.94 ± 0.03, Voting: 0.94 ± 0.02, $p < 0.001$). The voting method achieves the highest AUC value. The performance of multi-subject collaborative experiments can be simulated by regrouping the subjects into new groups with more members. In addition to the individual condition and the collaborative condition with two subjects, all 14 subjects are regrouped to groups with 3–14 subjects. Since the number of random combinations is too large to compute, the maximum number of random groups with a fixed number of subjects in one group is set to 100. The simulation results in **Figure 10** show that AUC increases significantly when the number of subjects is small, but AUC saturates (over 0.99 for both sessions) when the number of subjects reaches 5.

Cross-Session Collaborative Performance

The results of collaborative BCI performance in the cross-session condition are illustrated in **Figure 11**. The classification algorithm is SIM + HDCA (m = 21, according to the

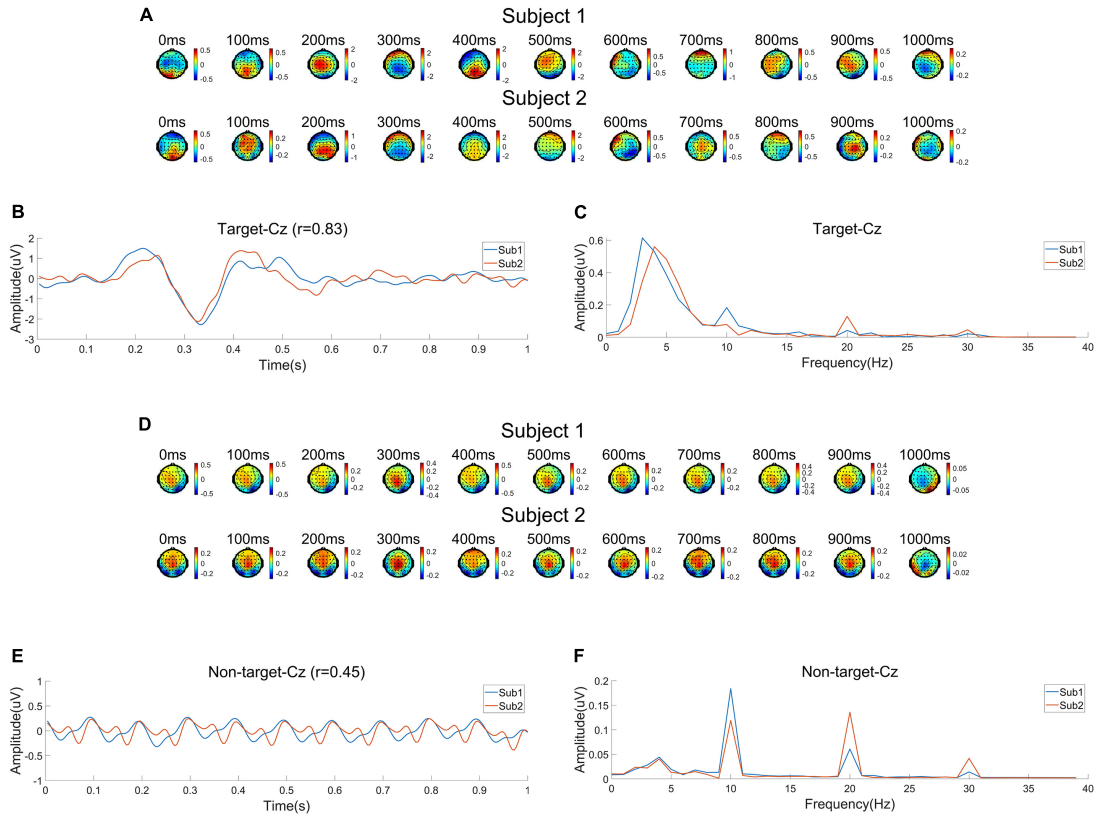


FIGURE 8 | Electroencephalogram (EEG) characteristics for a group of subjects (Subject 1 and Subject 2) on Day 1. **(A)** Time course of topographic maps of average event related potentials (ERPs) related to target images. **(B)** Average ERP waveform for targets at Cz. **(C)** Spectrum of ERPs for targets at Cz. **(D)** Time course of topographic maps of average EEG waveform related to non-target images. **(E)** Average EEG waveform for non-target images at Cz. **(F)** Spectrum of the average EEG waveform for non-target images at Cz.

individual results in **Figure 6**). As shown in **Figure 11**, all feature fusion algorithms can significantly improve the average performance of single subject (Single subject: 0.85 ± 0.06, ERP averaging: 0.87 ± 0.05, Feature concatenating: 0.90 ± 0.05, Voting: 0.90 ± 0.06, $p < 0.05$), and the voting method obtains the best performance. The ERP averaging method may be affected by the individual difference of amplitude and latency of ERPs, while the feature concatenating method increases the feature dimension which might increase the risk of overfitting. The voting method can avoid these problems by fusing the output scores together instead of fusing the EEG features. With the voting method, the cross-session collaborative performance is only slightly lower than the within-session collaborative performance and the difference is not significant (AUC: 0.90 ± 0.06 vs. 0.94 ± 0.02, $p > 0.05$). These results suggest that the cross-session method is efficient for the collaborative BCI. The simulated cross-session performance of a multi-subject BCI system is further shown in **Figure 10**. By increasing the number of subjects, the collaborative performance for the cross-session condition can also be significantly improved. The cross-session condition achieves similar performance to the within-session condition (Day 1) when the number of subjects increases to 10.

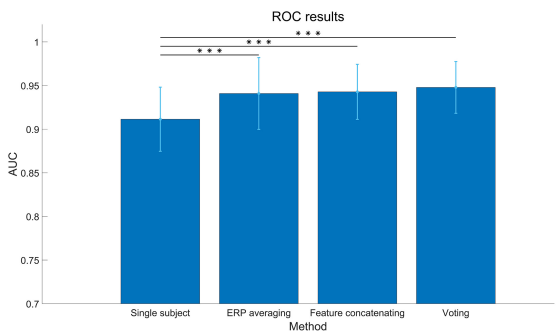


FIGURE 9 | Area under curve (AUC) of different feature fusion methods in the collaborative data analysis. The classification algorithm is SIM + hierarchical discriminant component analysis (HDCA; $m = 30$). The asterisks indicate that the performance of the fusion method is significantly higher than the single subject (***:$p < 0.001$). The error bars indicate standard deviations.

CONCLUSION AND DISCUSSION

This study presents a cross-session dataset of a collaborative RSVP-based BCI. The results illustrate the distinct spatial and

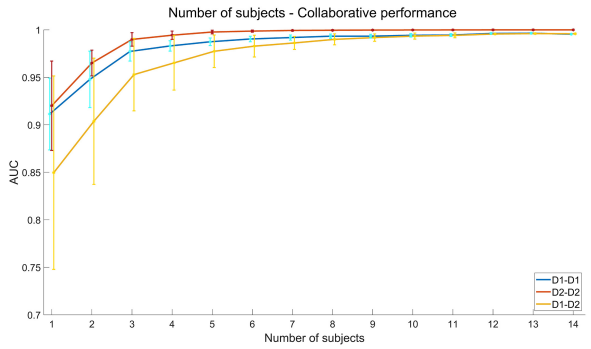


FIGURE 10 | Simulated performance of multi-subject collaborative data analysis. The *x* axis indicates the number of subjects in each group. The feature fusion method is voting. The error bars indicate standard deviations.

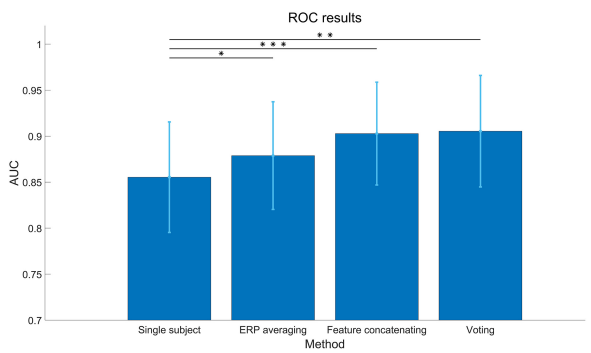


FIGURE 11 | Cross-session area under curve (AUC) of different feature fusion methods in the collaborative data analysis. The classification algorithm is SIM + hierarchical discriminant component analysis (HDCA; $m = 21$). The asterisks indicate that the performance of the fusion method is significantly higher than the single subject (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$). The error bars indicate standard deviations.

temporal features of ERPs related to target and non-target images. The comparison between different feature extraction and classification algorithms indicates that the combination of spatial filtering algorithms and HDCA can achieve good BCI performance in the individual condition, and the collaborative method can further improve the system performance by fusing information from multiple subjects. In the cross-session validation, the system performance can be optimized by selecting the number of components in the process by SNR maximizer for ERPs (SIM) algorithm. With the voting method, the cross-session collaborative performance is very close to the within-session collaborative performance (AUC: 0.90 vs 0.94). Although the cross-session AUC is still lower than the within-session AUC, the cross-session transfer can totally eliminate the system calibration procedure, which can substantially improve the practicality of the RSVP-based BCIs.

Since single-trial EEG data are recorded by multiple electrodes with various spatial and temporal features, suitable feature extraction and classification algorithms play important roles for ERP detection. In this paper, the SIM + HDCA algorithm achieves the best performance, but there is still room for improvement. First, the SSVEP component might contribute to ERP detection. As shown in the results, the single-trial EEG in RSVP tasks includes both ERPs and SSVEPs. However, the existing algorithms focus on the ERP components and ignore the SSVEP components in target detection. The difference of SSVEPs between target and non-target images requires further investigation by dissociating SSVEPs and ERPs (Zhang et al., 2018). Second, the number of components after spatial filtering was selected manually toward the highest AUC averaged across subjects and groups. The performance could be further improved by optimizing the number of components for each individual or group. In addition, the best number of components can be determined automatically by the algorithm toward a practical application.

The collaborative BCI method can be further improved by considering the following three directions. First, the feature fusion method can be improved by considering new features such as subject-to-subject synchronization or the response time of subjects (Poli et al., 2014; Valeriani et al., 2017). For instance, the EEG data of multiple subjects can be aligned by dynamic time warping (DTW) or canonical time warping (CTW) to synchronize the brain activities (Zhou and Torre, 2009). Second, the efficiency of the collaborative system can be optimized. For example, the BCI performance can be improved by collaborative paradigms with more subjects. However, a tradeoff between performance improvement and costs, which include equipment and labor costs in simultaneous EEG recording from multiple subjects, should be considered. When an individual subject achieves a high AUC value, the collaborative system can only obtain a minor improvement. Instead of the collaborative paradigm where the subjects perform the same detection tasks, another paradigm is to assign different tasks to each subject. This mode of division can improve the total efficiency of target detection tasks by reducing total time, but the individual performance remains the same. Third, CV can be combined with the RSVP-based BCI system. By optimally combining CV and HV, the system performance can be further improved (Sajda et al., 2010; Pohlmeyer et al., 2011).

In addition to the validation of collaborative and cross-session BCI performance in this study, this dataset can be used to study the following topics: (1) brain dynamics of ERPs and SSVEPs in the RSVP-based BCI paradigm, (2) data analysis algorithms for single-trial ERP detection, (3) data fusion methods for collaborative BCIs, and (4) transfer learning algorithms for the cross-session ERP-based BCIs.

DATA AVAILABILITY STATEMENT

The dataset analyzed in this study can be downloaded at https://doi.org/10.6084/m9.figshare.12824771.v1.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Tsinghua University. The

patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LZhe performed the data collection, data analysis, and wrote the manuscript. SS developed the experimental system and performed the data collection. HZ performed the data analysis. WP, HC, LZha, and XG revised the manuscript. YW supervised the study. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Key R&D Program of China under grant 2017YFA0205903, the National Natural Science Foundation of China under grant 61671424, and the Strategic Priority Research Program of Chinese Academy of Sciences under grant XDB32040200.

ACKNOWLEDGMENTS

The authors would like to thank the subjects for participating in the experiments, S. Zhang and B. Liu for offering advice on the manuscript, and C. Yang and H. Wu for building the website for data download.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2020.579469/full#supplementary-material

REFERENCES

Acqualagna, L., and Blankertz, B. (2013). Gaze-independent BCI-spelling using rapid serial visual presentation (RSVP). *Clin. Neurophysiol.* 124, 901–908. doi: 10.1016/j.clinph.2012.12.050

Acqualagna, L., Treder, M. S., Schreuder, M., and Blankertz, B. (2010). "A novel brain-computer interface based on the rapid serial visual presentation paradigm," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, Buenos Aires: IEEE, 2686–2689. doi: 10.1109/IEMBS.2010.5626548

Alpert, G. F., Manor, R., Spanier, A. B., Deouell, L. Y., and Geva, A. B. (2013). Spatiotemporal representations of rapid visual target detection: a single-trial EEG classification algorithm. *IEEE Trans. Biomed. Eng.* 61, 2290–2303. doi: 10.1109/TBME.2013.2289898

Bhattacharyya, S., Valeriani, D., Cinel, C., Citi, L., and Poli, R. (2019). "Collaborative brain-computer interfaces to enhance group decisions in an outpost surveillance task," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin: IEEE, 3099–3102. doi: 10.1109/EMBC.2019.8856309

Bigdely-Shamlo, N., Vankov, A., Ramirez, R. R., and Makeig, S. (2008). Brain activity-based image classification from rapid serial visual presentation. *IEEE Trans. Neural Syst. Rehabil. Eng.* 16, 432–441. doi: 10.1109/TNSRE.2008.2003381

Blankertz, B., Lemm, S., Treder, M., Haufe, S., and Müller, K. R. (2011). Single-trial analysis and classification of ERP components—a tutorial. *NeuroImage* 56, 814–825. doi: 10.1016/j.neuroimage.2010.06.048

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vis.* 10, 433–436. doi: 10.1163/156856897X00357

Broadbent, D. E., and Broadbent, M. H. (1987). From detection to identification: response to multiple targets in rapid serial visual presentation. *Percept. Psychophys.* 42, 105–113. doi: 10.3758/BF03210498

Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data Min.Knowl. Discov.*2, 121–167. doi: 10.1023/A:1009715923555

Cecotti, H., Eckstein, M. P., and Giesbrecht, B. (2014a). Single-trial classification of event-related potentials in rapid serial visual presentation tasks using supervised spatial filtering. *IEEE Trans. Neural Netw. Learn. Syst.* 25, 2030–2042. doi: 10.1109/TNNLS.2014.2302898

Cecotti, H., Eckstein, M. P., and Giesbrecht, B. (2014b). "Single-trial classification of neural responses evoked in rapid serial visual presentation: effects of stimulus onset asynchrony and stimulus repetition," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Chicago, IL: IEEE, 1282–1285. doi: 10.1109/EMBC.2014.6943832

Cecotti, H., and Graser, A. (2010). Convolutional neural networks for P300 detection with application to brain-computer interfaces. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 433–445. doi: 10.1109/TPAMI.2010.125

Cecotti, H., and Rivet, B. (2014). Subject combination and electrode selection in cooperative brain-computer interface based on event related potentials. *Brain Sci.* 4, 335–355. doi: 10.3390/brainsci4020335

Cecotti, H., Sato-Reinhold, J., Sy, J. L., Elliott, J. C., Eckstein, M. P., and Giesbrecht, B. (2011). "Impact of target probability on single-trial EEG target detection in a difficult rapid serial visual presentation task," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Boston, MA: IEEE, 6381–6384. doi: 10.1109/IEMBS.2011.6091575

Chun, M. M., and Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *J. Exp. Psychol. Human.* 21, 109–127. doi: 10.1037/0096-1523.21.1.109

Gao, S., Wang, Y., Gao, X., and Hong, B. (2014). Visual and auditory brain–computer interfaces. *IEEE Trans. Biomed. Eng.* 61, 1436–1447. doi: 10.1109/TBME.2014.2300164

Gerson, A. D., Parra, L. C., and Sajda, P. (2005). Cortical origins of response time variability during rapid discrimination of visual objects. *Neuroimage* 28, 342–353. doi: 10.1016/j.neuroimage.2005.06.026

Huang, Y., Erdogmus, D., Pavel, M., Mathan, S., and Hild, K. E. II (2011). A framework for rapid visual image search using single-trial brain evoked responses. *Neurocomputing* 74, 2041–2051. doi: 10.1016/j.neucom.2010.12.025

Hyvärinen, A., and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Netw.* 13, 411–430. doi: 10.1016/S0893-6080(00)00026-5

Jolicoeur, P. (1998). Modulation of the attentional blink by on-line response selection: evidence from speeded and unspeeded Task 1 decisions. *Mem. Cogn.* 26, 1014–1032. doi: 10.3758/BF03201180

Krauledat, M., Tangermann, M., Blankertz, B., and Müller, K. R. (2008). Towards zero training for brain-computer interfacing. *PLoS One* 3:2967. doi: 10.1371/journal.pone.0002967

Lawrence, D. H. (1971). Two studies of visual search for word targets with controlled rates of presentation. *Percept. Psychophys.* 10, 85–89. doi: 10.3758/BF03214320

Lees, S., Dayan, N., Cecotti, H., Mccullagh, P., Maguire, L., Lotte, F., et al. (2018). A review of rapid serial visual presentation-based brain–computer interfaces. *J. Neural. Eng.* 15:021001. doi: 10.1088/1741-2552/aa9817

Lees, S., McCullagh, P., Payne, P., Maguire, L., Lotte, F., and Coyle, D. (2019). Speed of rapid serial visual presentation of pictures, numbers and words affects event-related potential-based detection accuracy. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 113–122. doi: 10.1109/TNSRE.2019.2953975

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., et al. (2018). A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *J. Neural Eng.* 15:031005. doi: 10.1088/1741-2552/aab2f2/meta

Makeig, S., Bell, A. J., Jung, T. P., and Sejnowski, T. J. (1996). "Independent component analysis of electroencephalographic data," in *Advances in Neural Information Processing Systems 8*, eds D. Touretzky, M. Mozer, and M. Hasselmo (Cambridge MA: MIT Press), 145–151.

Mathan, S., Erdogmus, D., Huang, Y., Pavel, M., Ververs, P., Carciofini, J., et al. (2008). "Rapid image analysis using neural signals," in *CHI'08 Extended Abstracts on Human Factors in Computing Systems*, Florence, 3309–3314. doi: 10.1145/1358628.1358849

Matran-Fernandez, A., and Poli, R. (2017). Towards the automated localisation of targets in rapid image-sifting by collaborative brain-computer interfaces. *PLoS One* 12:e0178498. doi: 10.1371/journal.pone.0178498

Matran-Fernandez, A., Poli, R., and Cinel, C. (2013). "Collaborative brain-computer interfaces for the automatic classification of images," in *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, San Diego, CA: IEEE, 1096–1099. doi: 10.1109/NER.2013.6696128

Nakanishi, M., Wang, Y., Chen, X., Wang, Y. T., Gao, X., and Jung, T. P. (2018). Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis. *IEEE Trans. Biomed. Eng.* 65, 104–112. doi: 10.1109/TBME.2017.2694818

Oliva, A. (2005). "Gist of the scene," in *Neurobiology of Attention*, eds L. Itti, G. Rees, and J. Tsotsos (Cambridge, MA: Academic Press), 251–256. doi: 10.1016/B978-012375731-9/50045-8

Picton, T. W. (1992). The P300 wave of the human event-related potential. *J. Clin. Neurophysiol.* 9, 456–479. doi: 10.1097/00004691-199210000-00002

Pohlmeyer, E. A., Wang, J., Jangraw, D. C., Lou, B., Chang, S. F., and Sajda, P. (2011). Closing the loop in cortically-coupled computer vision: a brain–computer interface for searching image databases. *J. Neural. Eng.* 8:036025. doi: 10.1088/1741-2560/8/3/036025

Poli, R., Valeriani, D., and Cinel, C. (2014). Collaborative brain-computer interface for aiding decision-making. *PLoS One* 9:e102693.

Ramoser, H., Muller-Gerking, J., and Pfurtscheller, G. (2000). Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans. Rehabil.Eng.* 8, 441–446. doi: 10.1109/86.895946

Rivet, B., Souloumiac, A., Attina, V., and Gibert, G. (2009). xDAWN algorithm to enhance evoked potentials: application to brain–computer interface. *IEEE Trans. Biomed. Eng.* 56, 2035–2043. doi: 10.1109/TBME.2009.2012869

Rousselet, G. A., Thorpe, S. J., and Fabre-Thorpe, M. (2004). How parallel is visual processing in the ventral pathway? *Trends Cogn. Sci.* 8, 363–370. doi: 10.1016/j.tics.2004.06.003

Sajda, P., Gerson, A., and Parra, L. (2003). "High-throughput image search via single-trial event detection in a rapid serial visual presentation task," in *First International IEEE EMBS Conference on Neural Engineering, 2003. Conference Proceedings*, Capri Island: IEEE, 7–10. doi: 10.1109/CNE.2003.1196297

Sajda, P., Pohlmeyer, E., Wang, J., Parra, L. C., Christoforou, C., Dmochowski, J., et al. (2010). In a blink of an eye and a switch of a transistor: cortically coupled computer vision. *Proc. IEEE* 98, 462–478. doi: 10.1109/JPROC.2009.2038406

Serre, T., Oliva, A., and Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proc. Natl. Acad. Sci. U.S.A.* 104, 6424–6429. doi: 10.1073/pnas.0700622104

Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature* 381, 520–522. doi: 10.1038/381520a0

Touryan, J., Gibson, L., Horne, J. H., and Weber, P. (2011). Real-time measurement of face recognition in rapid serial visual presentation. *Front. Psychol.* 2:42. doi: 10.3389/fpsyg.2011.00042

Touyama, H. (2014). "A collaborative BCI system based on P300 signals as a new tool for life log indexing," in *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, San Diego, CA: IEEE, 2843–2846. doi: 10.1109/SMC.2014.6974360

Valeriani, D., Cinel, C., and Poli, R. (2017). Group augmentation in realistic visual-search decisions via a hybrid brain-computer interface. *Sci. Rep.* 7, 1–12. doi: 10.1038/s41598-017-08265-7

Valeriani, D., Poli, R., and Cinel, C. (2015). "A collaborative brain-computer Interface to improve human performance in a visual search task," in *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, (Montpellier: IEEE), 218–223. doi: 10.1109/NER.2015.7146599

Valeriani, D., Poli, R., and Cinel, C. (2016). Enhancement of group perception via a collaborative brain–computer interface. *IEEE Trans. Biomed. Eng.* 64, 1238–1248. doi: 10.1109/TBME.2016.2598875

Wang, Y., and Jung, T. P. (2011). A collaborative brain-computer interface for improving human performance. *PLoS One* 6:e0020422. doi: 10.1371/journal.pone.0020422

Wang, Y., Wang, Y. T., Jung, T. P., Gao, X., and Gao, S. (2011). "A collaborative brain-computer interface," in *2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)*, Vol. 1, Shanghai: IEEE, 580–583. doi: 10.1109/BMEI.2011.6098286

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain–computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 767–791. doi: 10.1016/S1388-2457(02)00057-3

Wu, W., and Gao, S. (2011). "Learning event-related potentials (ERPs) from multichannel EEG recordings: a spatio-temporal modeling framework with a fast estimation algorithm," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Boston, MA: IEEE, 6959–6962. doi: 10.1109/IEMBS.2011.6091759

Yuan, P., Wang, Y., Wu, W., Xu, H., Gao, X., and Gao, S. (2012). "Study on an online collaborative BCI to accelerate response to visual targets," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Diego, CA: IEEE, 1736–1739. doi: 10.1109/EMBC.2012.6346284

Zhang, S., Han, X., Chen, X., Wang, Y., Gao, S., and Gao, X. (2018). A study on dynamic model of steady-state visual evoked potentials. *J. Neural Eng.* 15:046010. doi: 10.1088/1741-2552/aabb82/meta

Zhao, H., Wang, Y., Sun, S., Pei, W., and Chen, H. (2019). "Obviating session-to-session variability in a rapid serial visual presentation-based brain–computer interface," in *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, San Francisco, CA: IEEE, 171–174. doi: 10.1109/NER.2019.8716892

Zhou, F., and Torre, F. (2009). Canonical time warping for alignment of human behavior. *Adv. Neural Informat. Process. Syst.* 22, 2286–2294.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

DATA REPORT
published: 26 October 2020
doi: 10.3389/fnins.2020.00919

HYGRIP: Full-Stack Characterization of Neurobehavioral Signals (fNIRS, EEG, EMG, Force, and Breathing) During a Bimanual Grip Force Control Task

Pablo Ortega[1,2], *Tong Zhao*[1] and *A. Aldo Faisal*[1,3,4*]

[1] Brain & Behavior Lab, Department of Computing and Department of Bioengineering, Imperial College London, London, United Kingdom, [2] EPSRC Centre for High Performance Embedded and Distributed Systems, Imperial College London, London, United Kingdom, [3] UKRI Centre in Artificial Intelligence for Healthcare, Imperial College London, London, United Kingdom, [4] Data Science Institute, London, United Kingdom

Keywords: electroencephalography, near-infrared spectroscopy, power-grip, non-invasive, brain-computer interface, sensor-fusion, data set, continuous decoding

OPEN ACCESS

Edited by:
Ian Daly,
University of Essex, United Kingdom

Reviewed by:
Yuriy Mishchenko,
Amazon, United States
Karen Moxon,
University of California, Davis, United States

***Correspondence:**
A. Aldo Faisal
aldo.faisal@imperial.ac.uk

Specialty section:
This article was submitted to Neural Technology, a section of the journal Frontiers in Neuroscience

Received: *31 May 2020*
Accepted: *10 August 2020*
Published: *26 October 2020*

Citation:
Ortega P, Zhao T and Faisal AA (2020) HYGRIP: Full-Stack Characterization of Neurobehavioral Signals (fNIRS, EEG, EMG, Force, and Breathing) During a Bimanual Grip Force Control Task. Front. Neurosci. 14:919. doi: 10.3389/fnins.2020.00919

Brain-computer interfaces (BCIs) have achieved important milestones in recent years, but the majority of breakthroughs in the continuous control of movement have focused on invasive neural interfaces with motor cortex or peripheral nerves. In contrast, non-invasive BCIs have primarily made progress in continuous decoding using event-related data, while the direct decoding of movement command or muscle force from brain data is an open challenge. Multi-modal signals from human cortex, obtained from mobile brain imaging that combines oxygenation and electrical neuronal signals, do not yet exploit their full potential due to the lack of computational techniques able to fuse and decode these hybrid measurements. To stimulate the research community and machine learning techniques closer to the state-of-the-art in artificial intelligence, we release herewith a holistic data set of hybrid non-invasive measures for continuous force decoding: the Hybrid Dynamic Grip (HYGRIP) data set. We aim to provide a complete data set that comprises the target force for the left/right-hand cortical brain signals in form of electroencephalography (EEG) with high temporal resolution and functional near-infrared spectroscopy (fNIRS), which captures in higher spatial resolution a BOLD-like cortical brain response, as well as the muscle activity (EMG) of the grip muscles, the force generated at the grip sensor (force), and confounding noise sources, such as breathing and eye movement activity during the task. In total, 14 right-handed subjects performed a uni-manual dynamic grip force task within 25–50% of each hand's maximum voluntary contraction. HYGRIP is intended as a benchmark with two open challenges and research questions for grip-force decoding. The first is the exploitation and fusion of data from brain signals spanning very different timescales, as EEG changes about three orders of magnitude faster than fNIRS. The second is the decoding of whole-brain signals associated with the use of each hand and the extent to which models share features for each hand or, conversely, are different for each hand. Our companion code makes the exploitation of the data readily available and accessible to researchers in the BCI, neurophysiology, and machine learning communities. HYGRIP can thus serve as a test bed for the development of BCI decoding algorithms and responses fusing multimodal brain signals. The resulting methods will help understand limitations and opportunities to benefit people in health and indirectly inform similar methods, answering the particular needs of people in disease.

INTRODUCTION

Brain-computer interfaces (BCIs) offer communication pathways for people with motor disorders to regain agency in their body and environment (Wolpaw et al., 2002). Since their first demonstration almost 50 years ago (Vidal, 1973, 1977; Wolpaw et al., 2000), BCIs have undergone a steady evolution. Invasive BCIs have achieved significant milestones in continuous signal read out from nervous system activity such as speech decoding (Guenther et al., 2009; Bocquelet et al., 2016; Anumanchipalli et al., 2019), robotic or own arm continuous control (Pfurtscheller et al., 2003; Hochberg et al., 2012), and even grip control with touch sense recovery (Ganzer et al., 2020). Non-invasive BCIs have also succeeded in the continuous control of trajectories after users learned to modulate event-related desynchronization (ERD) (Wolpaw and McFarland, 2004; Royer et al., 2010; Meng et al., 2016). However, while force is central to motor control (Westling and Johansson, 1984; Ostry and Feldman, 2003), its continuous non-invasive decoding is still challenging even in the offline case, and only modest accuracies have been reported using electroencephalography (EEG) (Paek et al., 2019). Previous attempts at decoding force from non-invasive measures have focused on the classification of discrete force variables using EEG (Jochumsen et al., 2013; Wang et al., 2017). In the hybrid case of recording cortical brain signals non-invasively, by combining EEG and functional near-infrared spectroscopy (fNIRS), Yin et al. (2015) showed that the combination of both measures increased the classification accuracies of different forces featured during imagined hand clenching by 1–5% compared to EEG or fNIRS alone. However, the lack of methods successfully integrating both measures in continuous decoding is still limiting the benefits of hybrid setups (Ahn and Jun, 2017).

We have shown that combining both multi-modal BCI (e.g., Thomik et al., 2013; Belić and Faisal, 2015; Xiloyannis et al., 2017) and the use of state-of-the-art machine learning—from introducing Deep Learning for EEG-BCI in 2015 (Walker et al., 2015) to data-efficient methods for BCI decoding that minimize the need for collecting data from individual end users (Xiloyannis et al., 2017; Ortega et al., 2018)—can help BCI research if data is collected with a machine learning use in mind. To stimulate the development of advanced multi-modal BCI techniques we present the Hybrid Dynamic Grip (HYGRIP) data set[1]. HYGRIP includes hybrid non-invasive and co-located brain activity measures as well as the hand contraction and muscular electrical behavioral activities during a hand-grip task with fast dynamics. The companion repository[2] digests the raw data into a format that makes it at a data readiness level suitable for immediate use by machine learning engineers (Lawrence, 2017) without having to go through a lengthy process of cleanup and reshaping of the data, which we believe will facilitate drawing in more data science and machine learning experts to the exciting problem of BCI.

[1]Raw data (2.8 h of data- worth time, 328 trials and 14 subjects) available at https://doi.org/10.6084/m9.figshare.12383639.v1.

[2]Companion code available at: https://gitlab.doc.ic.ac.uk/bbl/hygrip.git.

PARTICIPANTS

Fourteen ($N = 14$, anonymized IDs from A to N) healthy, right-handed volunteers participated in the production of this data set. Handedness was confirmed by the Edinburgh inventory (Oldfield, 1971) for all participants. None reported a history of neurological, cardio-respiratory, or physical disorders. Imperial College Research Ethics Committee approved all procedures, and all participants gave their written informed consent. The experiment complied with the Declaration of Helsinki for human experimentation and national and applicable international data protection rules.

MOTOR CONTROL TASK

The motor task consists of a left/right-hand grip, each hand being a different condition in the experiment. The task consisted of 10 consecutive contraction (1.55 s)/relaxation (0.55 s) periods that introduced rapid changes of force. Subjects were instructed and received visual feedback to exert forces in the 25–50% of their maximum voluntary contraction (MVC) following the pace of the 1.55 s contraction/0.55 s relaxation periods. The 25–50% MVC target range acted as a soft margin within which the subjects had to produce a contraction rather than a varying force they had to track. The MVC target range was implemented in this way to reduce the effect of visual feedback during the task that was provided through a computer screen for contractions out of the task range.

Due to the velocity of the contraction/relaxation periods, we do not consider each period a single trial but the consecutive 10 periods as a single trial of the task emphasizing the velocity of the execution. Note also that the much slower fNIRS signals need longer times to show a response, and using each single period independently of previous ones could hinder the resolution of the response. Participants were also instructed to prioritize the gripping pace rather than accurately matching the visual cues since the latter was too demanding for the levels of contraction required. Details follow.

Subjects sat in front of the computer screen with their arms relaxed and ergonomically hanging down, i.e., the arms were naturally straight downwards while holding the force transducer (**Figures 1A,B**). Subjects were instructed to keep this relaxed posture and reminded to maintain it throughout the experiment.

At the start of the experiment, subjects were asked to produce their maximum voluntary contraction (MVC) with each hand. The MVC was used to calibrate the feedback for each hand independently. To obtain reliable MVC estimates, MVC grips were repeated 10 times for 1 s with each hand and following paced auditory cues. The MVC was computed as the average of the maximum force across the 10 trials. For the experimental task, participants had to generate unimanual hand-grip contractions in the 25–50% MVC range during the 21 s gripping period. The contraction was computed as the ratio Force/MVC. The hand used for the grip varied randomly between trials with equal probability. The force transducer was handed by the experimenter to the subject within their immediate reach before each trial and removed after trial completion.

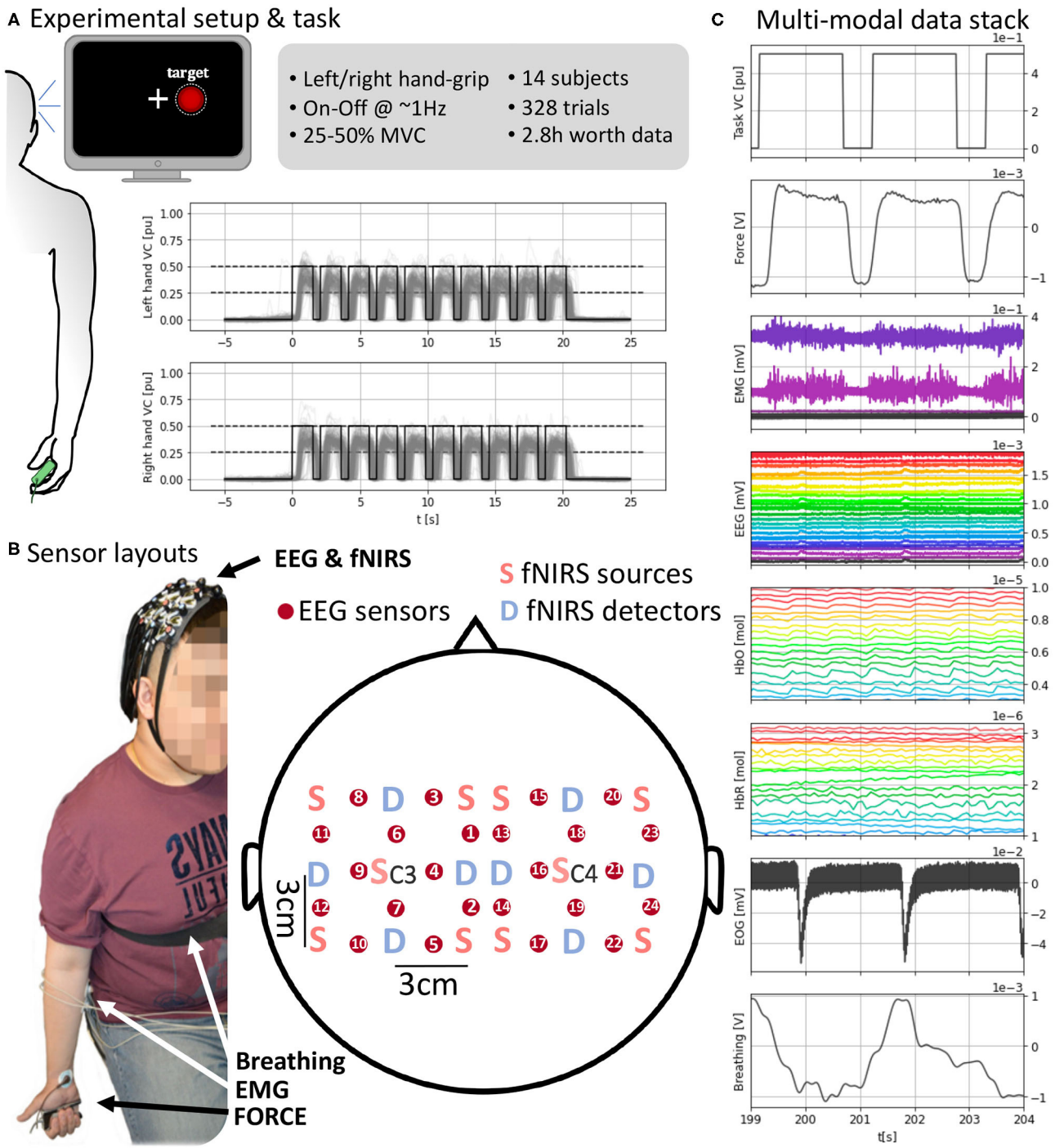


FIGURE 1 | (A) (Left) Experimental setup with visual feedback. (Top right) Task and data set details. (Bottom right) Task execution across subjects time-locked to the "Go" cue at $t = 0$ s. **(B)** Sensor layout and EEG and fNIRS sensor arrangement. fNIRS light sources and detectors are placed in a 5 × 5 symmetrical grid leaving a required 3 cm distance. EEG electrodes, red circles, are placed between each pair of sources and detectors overlaying the fNIRS sensing area so fNIRS and EEG measures are co-located. Each grid is centered around C3 and C4 for the respective hemisphere. **(C)** A subject's full stack of neurobehavioral data (5 s, selected channels per modality for readability). (Top-bottom) Force target for the right hand; force produced by subject's right hand; EMG channels from right forearm; EEG channels; HbO (oxyhemoglobin); HbR (deoxygenated hemoglobin); EOG with two eye blinks; breathing from chest strap stretch sensor. Units in the bottom x-axis and each corresponding y-axis.

A computer screen in front of the participants delivered the feedback at an approximate 60 Hz refreshing rate (**Figure 1B**). A cross indicating the center of the screen was used as fixation point during the experiment and subjects were instructed to stare at it to avoid eye movements. A white discontinuous circumference was used as a visual target and appeared at either side of the cross ipsilateral to the hand to be used in each trial. This visual target matched the 25–50% of the maximum voluntary contraction (MVC) of the subject for that hand and paced the contraction-relaxation periods. When the visual target appeared (shown for 1.55 s) the instruction was to contract and, when it disappeared (not shown for 0.55 s), to relax. A red filled circle indicated the real-time contraction level and subjects aimed to fill the discontinuous target circle with red. Visual feedback was only coupled to grips outside the desired range of contraction (25–50% MVC) making the red circle bigger or smaller than the target. Otherwise, the red circled filled the white discontinuous circumference within the target range of contraction.

Each trial of the task consisted of 10 consecutive contraction (1.55 s) and relaxation (0.55 s) periods (total, 21 s) with one hand (**Figure 1A**). The dynamic grip was executed with the hand indicated at the beginning of the trial by synthetic voice 2.5 s before the "Go" signal (used as the origin of time, $t = 0$ s, for each trial). All participants did a balanced amount of left- and right-hand trials of at least 10 (max. 13) trials per hand. We limited the number of trials to avoid effects of muscular fatigue given the relatively high contractions that the task demanded. Left and right-hand conditions were pseudorandomized across trials to avoid anticipation and interference between conditions. The refreshing rate of the visual feedback overlaid with the force produced in real-time. Each trial was followed by a randomized resting period uniformly distributed between 15 and 21 s, to avoid phasic constructive interference of systemic artifacts, e.g., Mayer waves, in the brain responses.

DATA COLLECTION PIPELINE AND METHODS

We recorded multiple signals representing brain activity, motor behavior, and confounds (**Figures 1C**, **2A–C**). The signals capturing brain activity consisted of electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS). Motor behavior was captured by the force sensor on which the subjects gripped and surface electromyography on both forearms. Potential confounds interfering (breathing and EOG) with EEG and fNIRS were also recorded. A total of three recording devices were used to record all the signals. fNIRS was recorded using a NIRScout system (NIRx Medizintechnik GmbH, Berlin, Germany). EEG, EMG, and EOG were recorded together with an ActiChamp amplifier (BrainProducts, Berlin, Germany). Force and breathing were recorded with a PowerLab 4/25T system (ADInstruments, Castle Hill, Australia). To synchronize the devices, the same computer used to present the task and visual feedback was used to send time-stamping signals to the three devices simultaneously at the beginning and end of the recording and every "Go" cue and were stored by each device in its time reference. The timestamps are used to locate the positions of the same event across different devices and align the measures to the events shown in the computer used to present the task. The sampling frequencies (12.5 Hz for fNIRS and 4 kHz for remaining measures) were selected so that they had a common divisor facilitating the resampling processes without the need to round up due to inexact divisors.

Brain Signals

All brain signals were non-invasively recorded. A custom 3D printed (formlabs Form2, Formlabs Inc., Somerville, MA 02143, USA) holder made of flexible resin (formlabs RS-F2-FLGR-02) was used to align the fNIRS and EEG sensors to approximately target similar cortical areas (**Figure 1B**). The sensor layout was configured to result in 12 hybrid EEG-fNIRS recording locations per hemisphere. These locations were homogeneously spread with a 3 cm separation creating a grid. Each hemispherical grid was centered around the corresponding 10-20 system C3 and C4 location.

fNIRS signals were recorded using a NIRScout system (NIRx Medizintechnik GmbH, Berlin, Germany). We used a total of 12 optodes per hemisphere (10 sources and eight detectors in total) sampling at 12.5 Hz. An optode is a source-detector pair 3 cm apart from each other (allowing light to reach an approximate 1.5 cm depth into the skull). fNIRS sources and sensors were laid out to result in 12 optodes. The sources and sensors were symmetrically laid around C3 and C4 positions according to the International 10-20 system leaving an inter-optode distance of 3 cm (**Figure 1B**). Two wavelengths ($wl_1 = 760$ nm, $wl_2 = 850$ nm) continuous functional near-infrared spectroscopy (fNIRS) was used to obtain optical absorption densities that were transformed to oxy-hemoglobin [HbO] and deoxy-hemoglobin concentrations [HbR] using the modified Beer-Lambert Law (Cope et al., 1988). The raw optical densities are also provided in the data set.

An ActiChamp amplifier (BrainProducts, Berlin, Germany) operating at 4 kHz (running software BrainVision, v1.20.0801) was used to record EEG. Twelve EEG sensors per hemisphere were placed in between each sensor-detector fNIRS pair overlaying the region measured by that optode (**Figure 1B**). The reference in our setup corresponds to the standard Cz 10-20 position (Nomenclature, 1991; Klem et al., 1999). The signals were down-pass filtered and downsampled to 1 kHz in the data set.

To enable EEG and fNIRS sensors to record cortical activity from the same cortical locations, we used a non-standard sensor arrangement covering the bilateral motor cortex (**Figure 1B**). We used a custom sensor holder 3D printed in flexible resin that for every recording channel allowed the EEG sensor to sit on top of the cortical area targeted by a corresponding fNIRS source-detector pair. Namely, for every fNIRS source-detector pair, an EEG electrode was placed in between. Each holder consisted of a 5×5 grid of circular holes whose centers were 1.5 cm apart allowing the required 3 cm separation between fNIRS source-detector pairs with an EEG sensor occupying a hole in between. The positions in the 5×5 grid marked in red in **Figure 1B** correspond to the physical location of EEG electrodes and

the approximate recording areas of EEG electrodes and fNIRS source-detector pairs. Physical locations of fNIRS sources and detectors are marked, respectively by "S" and "D" in **Figure 1B**. A total of 12 recording sites were used per hemisphere due to the limitation of space to fit the multi-modal sensors together. A holder was placed on each hemisphere and held in position using elastics. The disposition of every sensing point is symmetrical to the scalp mid-line, and each grid is centered in the corresponding C3 or C4 site depending on the side so that the central hole overlaid the respective central position. C3 and C4 were located per subject following the 10-20 standard (Nomenclature, 1991; Klem et al., 1999), i.e., at a 20% of the distance between the pre-auricular points passing over the top of the head from the mid-line. The spherical coordinates of the standard positions are provided in the data set.

Motor Signals

We recorded the grip force and muscular electrical activity to represent motor behavior during the task.

Bilateral bipolar surface electromyography (EMG) was recorded over the longitudinal axis (+ distal and − proximal) of the muscle belly of the *flexor digitorum superficialis* (4 kHz, on the Aux channels of the BrainVision ActiChamp) placed in the anterior and posterior forearm faces (**Figure 1B**). Before electrode placement, the skin was cleaned with abrasive pads and alcohol to eliminate dead skin cells and fat impact on electrical recording quality. EMG signals were down-pass filtered and downsampled to 1 kHz in the data set.

The dynamic gripping task was conducted using a continuously recorded grip force transducer (PowerLab 4/25T, ADInstruments, Castle Hill, Australia) sampling at 1 kHz. The signal was also used to provide real-time visual display feedback to the subject and the target force level they were asked to produce (**Figure 1A**). The Maximum Voluntary Contraction (MVC) force for each subject's hand was measured and computed using the same transducer. The recorded force signals, in Volts, were down-pass filtered and downsampled to 50 Hz in the published data set.

Recording of Potential Confounds (EOG and Breathing)

To complete the picture provided by the data, we recorded potential confounds in the brain and motor behavior signals of interest. The pulse and breathing rate have an impact at the body level on the concentration of hemoglobin and therefore can have an impact on brain and scalp levels of hemoglobin concentration. Whereas pulse is easily removed in the fNIRS analysis band (0.01–0.25 Hz), the breathing rate can overlap with it (Pinti et al., 2019). Sources of muscular electrical activity can spread to the EEG sensors and include eye muscles and skeletal muscles. Thus, we consider electrooculography (EOG), which also carries information on blinks, and breathing as pure confounds. However, EMG might also leak into the EEG sensors and can carry confound information at the same time it provides behavioral information.

Electrooculography (EOG)

Bipolar EOG was recorded on the vertical axis (top +, bottom −) of the right eye for all subjects (4 kHz, BrainVision ActiChamp, BrainProducts GmbH, Germany). The signal was down-pass filtered and downsampled to 1 kHz before being included in the data set. We note that the EOG of participant "I" is absent. Nonetheless, this participant was included as the impact of the EOG in the recording locations can be less severe than for frontal recording sites can be corrected using techniques like Independent Component Analysis (ICA) (Onton and Makeig, 2006).

Breathing

We captured breathing as the chest diametrical changes during inspiration and expiration. A variable resistor placed inside an elastic strap adjusted around the chest at the level of the Xiphoid process (**Figure 1B**) was used to record the expansion and contraction of the thoracic cage (1 kHz, PowerLab 4/25T, ADInstruments, Castle Hill, Australia). The signal was down-pass filtered and downsampled to 50 Hz before being included in the data set.

DATA SET OVERVIEW

The data set, provided as a single hard-disk file (HDF), has undergone very little processing to avoid biasing future analyses. Here, we make the raw data available and provide companion code that preprocesses the raw data into a readily usable data set (Data Readiness Level C). Preprocessing comprises down-sampling to reduce storage space and the formatting of data, recorded events, and other meta-data from different devices so that all data followed the same format regardless of their device origin. Further preprocessing can be directly applied using the `utils` python package provided, making the data readily available to exploit in popular python machine learning packages as `pytorch` and `tensorflow`.

Companion Code

The `utils` package only depends on the public python packages, `h5py`, `numpy`, `scipy`, `scikitlearn`, `matplotlib`, which need to be installed. The notebook `presentation.ipynb` contains a thorough explanation and examples of how the tools in `utils` can be used to process the data and depends on `jupyter`. A conda `environment.yml` file is provided with all dependencies to facilitate installation. All together make the data readily available to exploit, i.e., Data Readiness Level C (Lawrence, 2017).

Data Set Organization

The data set file structure follows a tree-like organization in three levels. In the first and third levels, the data set contains meta-data in *string* format that can be accessed via the attributes of the level. The first level is the data set itself and the shared attributes across subjects measures, e.g., sampling frequencies and units, and other information such as the channel grid disposition and a template of hybrid sensor spherical coordinates over the scalp. In the second level, the data set is organized in one group per subject

indexed by their anonymized ID (i.e., 14 groups with keys `A` to `N`) and contain no attributes. In the third level, each group contains a subgroup for each measure (e.g., keys `frc` for force and `eeg` for EEG) containing the data in *numeric* format and an attribute called `events` containing the times at which a timestamp was received during the recording (e.g., "relax," "left-hand," "right-hand") and the "begin" and "end" timestamps indicating the beginning and end of the recording session, also *numeric*. In particular, the `events` group contains a second *numeric* attribute `MVC` containing the maximum voluntary contraction value for each hand.

Data Validation Preprocessing Pipeline

The following preprocessing was applied to the raw signals in the data set to obtain the brief analysis in **Figure 2**, which can be reproduced using the companion notebook. After preprocessing all signals, epochs were extracted from 5 s before the "Go" instruction to 25 s after.

fNIRS

The optical intensity, $\hat{\mathrm{I}}_{ij}^{\lambda}$, for each wavelength, λ, was low-pass filtered below 0.25 Hz with a 7th order elliptical filter. Changes in optical densities per wavelength, $\Delta\mathrm{OD}_{ij}^{\lambda}(t)$, were obtained using

$$\Delta\mathrm{OD}_{ij}^{\lambda}(t) = -\log\left(\hat{\mathrm{I}}_{ij}^{\lambda}(t) \,/\, \bar{\mathrm{I}}_{ij}^{\lambda}\right) \tag{1}$$

with i and j the indices of valid sensor-detector pairs respectively, t the time and $\bar{\mathrm{I}}_{ij}^{\lambda}$ the average of the optical intensity 1 s prior to the "Go" instruction. Oxygenated and deoxygenated hemoglobin concentration changes, ΔHbO and ΔHbR, respectively, were computed solving the modified Beer-Lambert law (Cope et al., 1988),

$$\Delta\mathrm{OD}_{ij}^{\lambda} = \mathrm{L}_{ij}^{\lambda}\mathrm{DPF}^{\lambda}(\epsilon_{\mathrm{HbR}}^{\lambda}\Delta\mathrm{HbR} + \epsilon_{\mathrm{HbO}}^{\lambda}\Delta\mathrm{HbO}) \tag{2}$$

with DPF^{λ}, the dimensionless *differential path-length factor* accounting for the reduction in intensity due to scattering tissues ($\mathrm{DPF}^{760} = 5.98$ and $\mathrm{DPF}^{850} = 7.54$); $\epsilon_{\mathrm{Hb}}^{\lambda}$, the *molar extinction coefficient* for each hemoglobin and wavelength in $\mathrm{mol}^{-1}\mathrm{cm}^{-1}$ accounting for the absorption of light ($\epsilon_{\mathrm{HbO}}^{760} = 1486.6$, $\epsilon_{\mathrm{HbR}}^{760} = 3843.7$, $\epsilon_{\mathrm{HbO}}^{850} = 2526.4$, and $\epsilon_{\mathrm{HbR}}^{850} = 1798.6$); and L the source-detector distance in cm ($\mathrm{L} = 3\mathrm{cm}$). After this preprocessing, the average and the standard error of the mean across subjects for each hand condition and hemoglobin type were plotted in the corresponding position of the 2D layout (**Figure 2A**). We can observe an increase and decrease of HbO with the task onset ($t = 0$ s) and end ($t = 21$ s), respectively. We can also identify several peaks in the average response which might be a result of the on-off dynamics of the task which might introduce small variations on the global trend of Hb variations. Opposite changes can be observed for HbR at a smaller scale.

EEG

EEG was first downsampled to 250 Hz (with anti-aliasing down-pass filtering). Notch filters were applied at the mains (50 Hz) and fNIRS (12.5 Hz) frequencies and their harmonics. EEG was finally high-pass filtered above 1 Hz using a 5th order Butterworth filter. ICA was used in two stages to remove components correlated first with EOG and second with EMG. ICA related preprocessing only affected the signals used to compute the mixing matrix, which was then applied to the data going through the main EEG pipeline. For EOG, both the EOG and EEG were downsampled to 25 Hz. A maximum of 1 independent component correlated above 0.3 (in absolute values) with the EOG was rejected. For EMG, both the EMG and EEG were downsampled to 125 Hz. The rejection of components was stricter to ensure EMG was not contaminating the data. One component was rejected whenever its correlation magnitude with any of the recorded EMG channels was $>10^{-4}$. **Figure 2B** presents the averaged spectrogram across subjects for the right-hand condition (the left-hand condition can be found in the notebook) for the mu band (8–13 Hz). Interestingly, the on-off nature of the task might be introducing periodic variations of power in the mu band due to desynchronization (Pfurtscheller et al., 2006).

EMG

The EMG was also first downsampled to 250 Hz (with anti-aliasing down-pass filtering), and it was then high-pass filtered with a 17th order Butterworth filter of above 110 Hz. To generate **Figure 2C** we computed the Hilbert envelope of the signal and used it to obtain decibels of power density referred to the mean power of the signal during the epoch. We finally averaged these power densities across subjects per hand condition and arm location of the electrodes. The active EMG (i.e., those corresponding to the arm used during each hand condition) carry a similar amount of power density for each hand condition. The passive electrodes have a much flatter amount of power density during the task and instead have clear peaks at the beginning and end of the trials when the subjects were allowed to relax.

Force

The force signal was band-pass filtered between 10^{-4} and 9 Hz (second order elliptical filter). Once epochs were extracted, it was again high-pass filtered above 10^{-3} Hz to remove any remaining offset. These low high-pass frequencies were selected to preserve the squared shape of the forces, which are very rich in low frequencies. Once the offset was removed, voluntary contraction values were obtained by dividing the resulting forces by the maximum voluntary contraction force recorded at the beginning of the trial. **Figure 1A** shows the gathered trials for all subjects per hand condition. Subjects mostly engaged with the task in timing and contraction values with the left condition presenting slightly more overshoots. Although the task is conceptually simple, the provision of only partial visual feedback and its fast on-off nature contributed to higher variability in the behavior within the desired levels of contraction. We consider these aspects to be more representative of natural force applications where feedback is more proprioceptive and changes in force can be fast and span a wider range than discrete target levels.

Other Signals (EOG and Breathing)

EOG was downsampled (with prior anti-aliasing filtering) to 50 Hz. Then filtered using an 8th order high-pass filter above

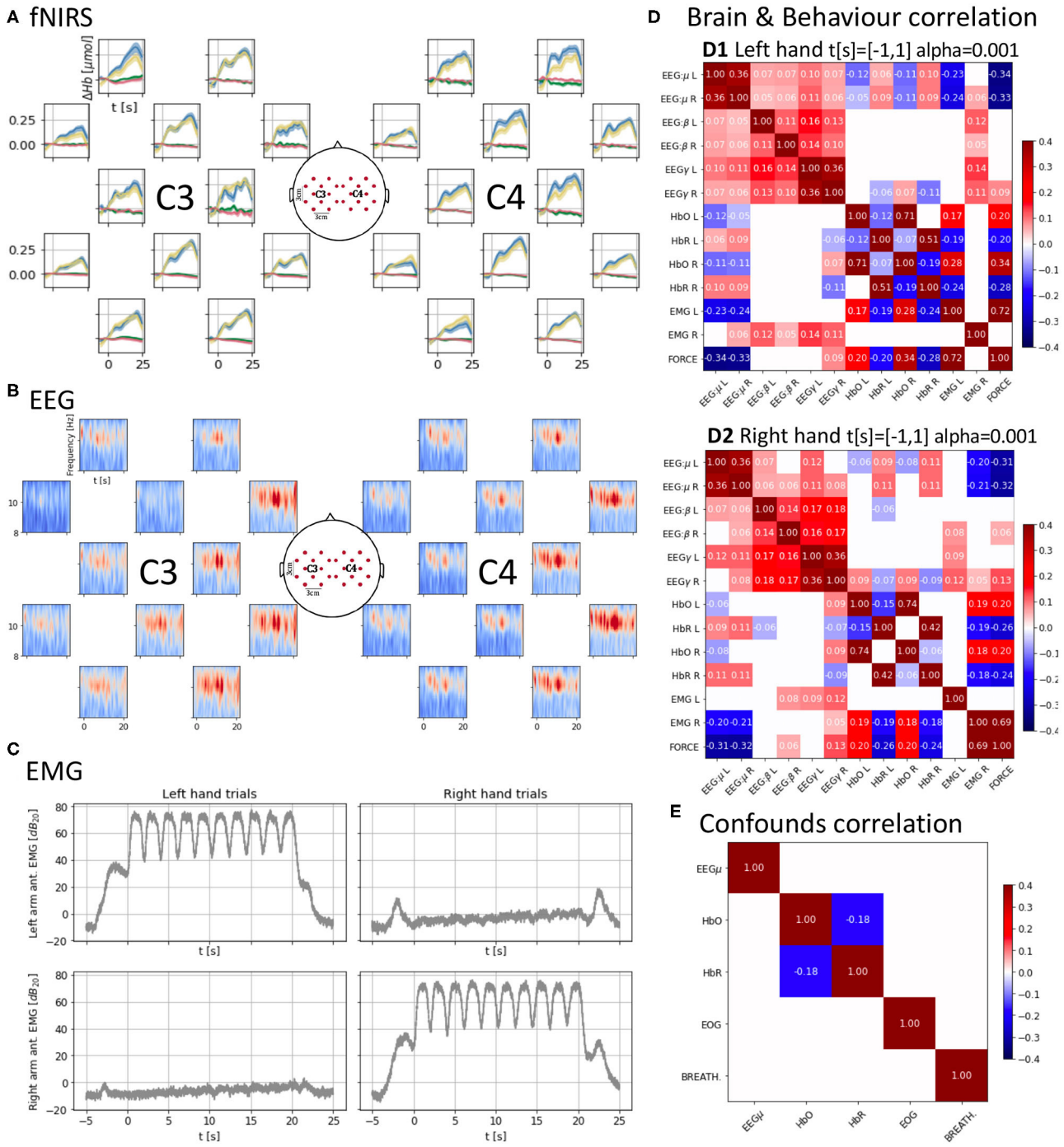


FIGURE 2 | (A) Cross-subject average of HbO and HbR changes across trials, time-locked to the "Go" cue ($t = 0$ s) of each trial. Each trial consists of 10 contraction/relaxation periods (21 s of activity starting at the "Go" cue). HbO increases during the task and HbR decreases at a smaller scale, both start returning to baseline after the 21 s of activity. Units in top-left. **(B)** Cross-subject mu-band spectrogram averaged across right-hand trials as in **(B)**. Periodic desynchronizations can be observed. Units in top-left. **(C)** Averaged EMG spectral density across trials as in **(A,B)** showing similar power density for the active muscles controlling the hands. **(D)** Brain and behavior correlation matrix computed on the force onset (−1 s to 1 s around the "Go" cue) showing only significant ($p < 0.001$) correlations. It shows a contralateral change in correlation values between HbO and HbR and force and the time-locked mu desynchronization event represented as the negative correlation between force and the EEG mu power band. **(E)** Confounds correlation matrix showing only significant ($p < 0.001$) correlations computed along the task ($t = [-1, 20]$ s around the "Go" cue) for the right hand condition. Correlations between brain signals and confounds are non-significant ($\alpha = 0.001$).

1 Hz. Breathing was also low-pass filtered with a 6th order Butterworth filter below 0.25 Hz.

Correlations Across Multi-Modal Signals and Validation

After preprocessing, signals were correlated to characterize a "brain and behavior" or neurobehavioral correlation (EEG, fNIRS, EMG, and force in **Figure 2D**) and a "confounds" correlation structure (EEG, fNIRS, EOG, and breathing in **Figure 2E**). EEG and EMG underwent additional filtering and spectral density computation similarly to that used for the EMG plots. For EEG the mu (8–12 Hz), beta (12–30 Hz), and gamma (60–125 Hz) bands spectral densities were extracted. Then, signals were downsampled to the lowest sampling frequency in the data set, i.e., 12.5 Hz, and cropped between -1 and 1 s ("Go" cue at $t = 0$ s) to focus on the onset of motor activity in the case of the "brain and behavior" correlation and between -1 and 21 s for the "confounds." The signals were finally normalized. For each hand condition, the "brain and behavior" correlation (**Figure 2D**) was computed over the appended observations corresponding to the $t = [-1, 1]$ s crops for all combinations of the left and right hemisphere mu, beta, and gamma EEG power bands and HbO and HbR, the right- and left-arm EMG and force. For the brain signals, channels 4 and 16 (**Figure 1B**) were selected as representative of the corresponding hemisphere activity. A similar process was used for the "confounds" correlation for the $t = [-1, 20]$ s. A significance level $\alpha = 0.001$ is set and only significant correlation values are shown.

There is a very small correlation between brain signals and confounds (< 0.03 in absolute values, not shown due to lack of significance $\alpha = 0.001$, **Figure 2E**). This suggests that the confounds do not interfere with the recorded signals after applying the standard preprocessing pipeline to the raw data.

The key observations of our data can be found for the "brain and behavior" correlation (**Figure 2D**). For considering the interaction between brain signals (fNIRS and EEG) and task measure (force) we need to bear in mind that the time scales of fNIRS and EEG are very different (seconds vs. milliseconds). In the EEG domain, it is known that motor activity onsets are reflected in EEG power features. In our data, we observe significant strong anti-correlation ($r \approx -0.3$, $p < 0.001$) between the EEG power in the mu band and the force which indicates that passing from a resting state (high mu power) to motor activity (low mu power) is properly captured by these spectral features when we look at a time window from -1 to 1 s around the Go cue. This decrease in power, also known as *mu event-related desynchronization* (mu-ERD), is due to the desynchronization of neuronal activity (Pfurtscheller et al., 2006), and it shows that the EEG is aligned with the force, helping to further confirm the validity of our data set.

However, when the correlation is computed focusing on the 20 s of the task (from 0 s before the Go cue to 20 s after), the EEG mu-ERD is not significant ($p > 0.001$). Only the beta band (12–30 Hz), known to be synchronized with motor activity (Kristeva-Feige et al., 2002), appears with a low level positive correlation ($r \approx 0.04 - 0.05$, $p < 0.001$). The lack of mu-ERD and force correlation during the 20 s of continuous contraction/relaxations might be a consequence of the velocity at which the sequential contractions/relaxations were executed, not leaving enough time to the motor cortex to reach a synchronized equilibrium state before it was desynchronized again. Furthermore, this can also indicate that mu frequencies (8–12 Hz) are not fast enough to track this kind of subtle phase changes and a justification to develop more precise algorithms or feature extractors as suggested by Paek et al. (2019).

In the fNIRS domain ± 1 s around the "Go" cue, we also observe a typical HbO/HbR anticorrelation ($r \approx 0.15$, $p < 0.001$) (Jasdzewski et al., 2003; Huppert et al., 2006) in the structure with higher magnitudes present for the contralateral hemisphere to the hand used. The HbO/HbR anticorrelation is stronger in the right hemisphere for the left hand although also present in the left hemisphere, and it is stronger in the right hemisphere. This suggests that the dominant right hand may engage the left hemisphere while the non-dominant left hand engages both hemispheres with a preference for the contralateral one. HbO and HbR also show significant correlation with the force ($r \approx 0.2 - 0.3$, $p < 0.001$). In particular, HbR, which is more specific than HbO (Hirth et al., 1997), also shows higher anticorrelation with the force for the contralateral hemisphere.

Finally, there is a strong significant correlation ($r \approx 0.7$, $p < 0.001$) between the EMG power envelopes of the active hands and the force which supports the synchronization of the different devices used to record these measures.

This brief analysis aims to validate the data set and present some of its features. We encourage the community to develop algorithms to better understand the rich temporal relationships between brain signals spanning very different time scales and physical origins present in the data set with BCI and neuroimaging purposes.

DATA AVAILABILITY STATEMENT

The HYGRIP data set can be downloaded from https://doi.org/10.6084/m9.figshare.12383639.v1 and the companion code can be cloned from https://gitlab.doc.ic.ac.uk/bbl/hygrip.git.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Imperial College London Research Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AF conceived the project. PO and AF designed the experiment, analyzed the data, and edited the manuscript. PO and TZ performed the data recordings. PO wrote the companion code

and drafted the paper. All authors contributed to the article and approved the submitted version.

FUNDING

PO was supported by a Ph.D. training scholarship by the EPSRC Centre for Doctoral Training HiPEDS (EP/L016796/1). The fNIRS system was available through an EPSRC capital equipment support grant to AF.

ACKNOWLEDGMENTS

We thank the volunteers for their time and cooperation and the reviewers for their valuable feedback.

REFERENCES

Ahn, S. and Jun, S. C. (2017). Multi-modal integration of EEG-fNIRS for brain-computer interfaces-current limitations and future directions. *Front. Hum. Neurosci.* 11:503. doi: 10.3389/fnhum.2017.00503

Anumanchipalli, G. K., Chartier, J., and Chang, E. F. (2019). Speech synthesis from neural decoding of spoken sentences. *Nature* 568, 493–498. doi: 10.1038/s41586-019-1119-1

Belić, J. J. and Faisal, A. A. (2015). Decoding of human hand actions to handle missing limbs in neuroprosthetics. *Front. Comput. Neurosci.* 9:27. doi: 10.3389/fncom.2015.00027

Bocquelet, F., Hueber, T., Girin, L., Savariaux, C., and Yvert, B. (2016). Real-time control of an articulatory-based speech synthesizer for brain computer interfaces. *PLoS Comput. Biol.* 12:e1005119. doi: 10.1371/journal.pcbi.1005119

Cope, M., Delpy, D., Reynolds, E., Wray, S., Wyatt, J., and Van der Zee, P. (1988). "Methods of quantitating cerebral near infrared spectroscopy data," in *Oxygen Transport to Tissue X*, eds M. Mochizuki, C. R. Honig, T. Koyama, T. K. Goldstick, and D. F. Bruley (New York, NY: Springer), 183–189. doi: 10.1007/978-1-4615-9510-6_21

Ganzer, P. D., Colachis, S. C. IV, Schwemmer, M. A., Friedenberg, D. A., Dunlap, C. F., Swiftney, C. E., et al. (2020). Restoring the sense of touch using a sensorimotor demultiplexing neural interface. *Cell* 181, 763–773.e12. doi: 10.1016/j.cell.2020.03.054

Guenther, F. H., Brumberg, J. S., Wright, E. J., Nieto-Castanon, A., Tourville, J. A., Panko, M., et al. (2009). A wireless brain-machine interface for real-time speech synthesis. *PLoS ONE* 4:e8218. doi: 10.1371/journal.pone.0008218

Hirth, C., Obrig, H., Valdueza, J., Dirnagl, U., and Villringer, A. (1997). "Simultaneous assessment of cerebral oxygenation and hemodynamics during a motor task," in *Oxygen Transport to Tissue XVIII*, eds E. M. Nemoto and C. J. LaManna (New York, NY: Springer), 461–469. doi: 10.1007/978-1-4615-5865-1_59

Hochberg, L. R., Bacher, D., Jarosiewicz, B., Masse, N. Y., Simeral, J. D., Vogel, J., et al. (2012). Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature* 485, 372–375. doi: 10.1038/nature11076

Huppert, T. J., Hoge, R. D., Diamond, S. G., Franceschini, M. A., and Boas, D. A. (2006). A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *NeuroImage* 29, 368–382. doi: 10.1016/j.neuroimage.2005.08.065

Jasdzewski, G., Strangman, G., Wagner, J., Kwong, K., Poldrack, R., and Boas, D. (2003). Differences in the hemodynamic response to event-related motor and visual paradigms as measured by near-infrared spectroscopy. *Neuroimage* 20, 479–488. doi: 10.1016/S1053-8119(03)00311-2

Jochumsen, M., Niazi, I. K., Mrachacz-Kersting, N., Farina, D., and Dremstrup, K. (2013). Detection and classification of movement-related cortical potentials associated with task force and speed. *J. Neural Eng.* 10:056015. doi: 10.1088/1741-2560/10/5/056015

Klem, G. H., Lüders, H. O., Jasper, H., and Elger, C. (1999). The ten-twenty electrode system of the international federation. *Electroencephalogr. Clin. Neurophysiol.* 52, 3–6.

Kristeva-Feige, R., Fritsch, C., Timmer, J., and Lücking, C.-H. (2002). Effects of attention and precision of exerted force on beta range EEG-EMG synchronization during a maintained motor contraction task. *Clin. Neurophysiol.* 113, 124–131. doi: 10.1016/S1388-2457(01)00722-2

Lawrence, N. D. (2017). Data readiness levels. *arXiv [preprint]* arXiv:1705.02245.

Meng, J., Zhang, S., Bekyo, A., Olsoe, J., Baxter, B., and He, B. (2016). Noninvasive electroencephalogram based control of a robotic arm for reach and grasp tasks. *Sci. Rep.* 6:38565. doi: 10.1038/srep38565

Nomenclature, S. E. P. (1991). American electroencephalographic society guidelines for. *J. Clin. Neurophysiol.* 8, 200–202. doi: 10.1097/00004691-199104000-00007

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Onton, J., and Makeig, S. (2006). Information-based modeling of event-related brain dynamics. *Prog. Brain Res.* 159, 99–120. doi: 10.1016/S0079-6123(06)59007-7

Ortega, P., Colas, C., and Faisal, A. A. (2018). "Compact convolutional neural networks for multi-class, personalised, closed-loop EEG-BCI," in *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)* (Enschede), 136–141. doi: 10.1109/BIOROB.2018.8487644

Ostry, D. J., and Feldman, A. G. (2003). A critical evaluation of the force control hypothesis in motor control. *Exp. Brain Res.* 153, 275–288. doi: 10.1007/s00221-003-1624-0

Paek, A. Y., Gailey, A., Parikh, P. J., Santello, M., and Contreras-Vidal, J. L. (2019). Regression-based reconstruction of human grip force trajectories with noninvasive scalp electroencephalography. *J. Neural Eng.* 16:066030. doi: 10.1088/1741-2552/ab4063

Pfurtscheller, G., Brunner, C., Schlögl, A., and Da Silva, F. L. (2006). Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks. *NeuroImage* 31, 153–159. doi: 10.1016/j.neuroimage.2005.12.003

Pfurtscheller, G., Müller, G. R., Pfurtscheller, J., Gerner, H. J., and Rupp, R. (2003). "Thought"-control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia. *Neurosci. Lett.* 351, 33–36. doi: 10.1016/S0304-3940(03)00947-9

Pinti, P., Scholkmann, F., Hamilton, A., Burgess, P., and Tachtsidis, I. (2019). Current status and issues regarding pre-processing of fNIRS neuroimaging data: an investigation of diverse signal filtering methods within a general linear model framework. *Front. Hum. Neurosci.* 12:505. doi: 10.3389/fnhum.2018.00505

Royer, A. S., Doud, A. J., Rose, M. L., and He, B. (2010). EEG control of a virtual helicopter in 3-dimensional space using intelligent control strategies. *IEEE Trans. Neural Syst. Rehabil. Eng.* 18, 581–589. doi: 10.1109/TNSRE.2010.2077654

Thomik, A. A., Haber, D., and Faisal, A. A. (2013). "Real-time movement prediction for improved control of neuroprosthetic devices," in *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)* (San Diego, CA), 625–628. doi: 10.1109/NER.2013.6696012

Vidal, J. J. (1973). Toward direct brain-computer communication. *Annu. Rev. Biophys. Bioeng.* 2, 157–180. doi: 10.1146/annurev.bb.02.060173.001105

Vidal, J. J. (1977). Real-time detection of brain events in EEG. *Proc. IEEE* 65, 633–641. doi: 10.1109/PROC.1977.10542

Walker, I., Deisenroth, M., and Faisal, A. (2015). *Deep Convolutional Neural Networks for Brain Computer Interface Using Motor Imagery*. Imperial College of Science, Technology and Medicine, Department of Computing, 68.

Wang, K., Wang, Z., Guo, Y., He, F., Qi, H., Xu, M., et al. (2017). A brain-computer interface driven by imagining different force loads on a single hand: an online feasibility study. *J. Neuroeng. Rehabil.* 14:93. doi: 10.1186/s12984-017-0307-1

Westling, G., and Johansson, R. (1984). Factors influencing the force control during precision grip. *Exp. Brain Res.* 53, 277–284. doi: 10.1007/BF00238156

Wolpaw, J. R., Birbaumer, N., Heetderks, W. J., McFarland, D. J., Peckham, P. H., Schalk, G., et al. (2000). Brain-computer interface technology: a review of the first international meeting. *IEEE Trans. Rehabil. Eng.* 8, 164–173. doi: 10.1109/TRE.2000.847807

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 767–791. doi: 10.1016/S1388-2457(02)00057-3

Wolpaw, J. R., and McFarland, D. J. (2004). Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proc. Natl. Acad. Sci. U.S.A.* 101, 17849–17854. doi: 10.1073/pnas.0403504101

Xiloyannis, M., Gavriel, C., Thomik, A. A., and Faisal, A. A. (2017). Gaussian process autoregression for simultaneous proportional multi-modal prosthetic control with natural hand kinematics. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 1785–1801. doi: 10.1109/TNSRE.2017.2699598

Yin, X., Xu, B., Jiang, C., Fu, Y., Wang, Z., Li, H., et al. (2015). A hybrid BCI based on EEG and fNIRS signals improves the performance of decoding motor imagery of both force and speed of hand clenching. *J. Neural Eng.* 12:036004. doi: 10.1088/1741-2560/12/3/036004

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 27 November 2020
doi: 10.3389/fnins.2020.589659

Evaluation of a P300-Based Brain-Machine Interface for a Robotic Hand-Orthosis Control

Jonathan Delijorge[1]*, *Omar Mendoza-Montoya*[1], *Jose L. Gordillo*[1], *Ricardo Caraza*[2], *Hector R. Martinez*[2] *and Javier M. Antelis*[1]

[1] Tecnologico de Monterrey, Escuela de Ingeniería y Ciencias, Monterrey, Mexico, [2] Tecnologico de Monterrey, Escuela de Medicina y Ciencias de la Salud, Monterrey, Mexico

This work presents the design, implementation, and evaluation of a P300-based brain-machine interface (BMI) developed to control a robotic hand-orthosis. The purpose of this system is to assist patients with amyotrophic lateral sclerosis (ALS) who cannot open and close their hands by themselves. The user of this interface can select one of six targets, which represent the flexion-extension of one finger independently or the movement of the five fingers simultaneously. We tested offline and online our BMI on eighteen healthy subjects (HS) and eight ALS patients. In the offline test, we used the calibration data of each participant recorded in the experimental sessions to estimate the accuracy of the BMI to classify correctly single epochs as target or non-target trials. On average, the system accuracy was 78.7% for target epochs and 85.7% for non-target trials. Additionally, we observed significant P300 responses in the calibration recordings of all the participants, including the ALS patients. For the BMI online test, each subject performed from 6 to 36 attempts of target selections using the interface. In this case, around 46% of the participants obtained 100% of accuracy, and the average online accuracy was 89.83%. The maximum information transfer rate (ITR) observed in the experiments was 52.83 bit/min, whereas that the average ITR was 18.13 bit/min. The contributions of this work are the following. First, we report the development and evaluation of a mind-controlled robotic hand-orthosis for patients with ALS. To our knowledge, this BMI is one of the first P300-based assistive robotic devices with multiple targets evaluated on people with ALS. Second, we provide a database with calibration data and online EEG recordings obtained in the evaluation of our BMI. This data is useful to develop and compare other BMI systems and test the processing pipelines of similar applications.

Keywords: brain-machine interface, electroencephalography, evoked potentials, P300, amyotrophic lateral sclerosis, signal processing, artificial intelligence, hand-orthosis

OPEN ACCESS

Edited by:
Davide Valeriani, Harvard Medical School, United States

Reviewed by:
Rupert Ortner, g.tec Medical Engineering Spain S.L., Spain
Floriana Pichiorri, Santa Lucia Foundation (IRCCS), Italy

***Correspondence:**
Jonathan Delijorge
delijorge@tec.mx

Specialty section:
This article was submitted to Neural Technology, a section of the journal Frontiers in Neuroscience

Received: *31 July 2020*
Accepted: *22 October 2020*
Published: *27 November 2020*

Citation:
Delijorge J, Mendoza-Montoya O, Gordillo JL, Caraza R, Martinez HR and Antelis JM (2020) Evaluation of a P300-Based Brain-Machine Interface for a Robotic Hand-Orthosis Control. Front. Neurosci. 14:589659. doi: 10.3389/fnins.2020.589659

1. INTRODUCTION

Since the early developments of BMIs, one of the most promising applications of this technology is the use of neuroprosthetic devices to assist people with reduced mobility. There is a consensus among researchers of this area that BMIs may significantly improve the lives of patients who suffer neuromuscular disorders such as ALS. Even so, despite all the efforts in the last three decades to

design and implement reliable BMI systems, the goal of developing functional neuroprostheses has not been reached yet. Researchers and engineers must solve many technical and practical problems before bringing this technology into everyday life. Some open issues concerning the development of robust brain-controlled applications are the ability of the system to interpret the user's intentions accurately, the time to process and analyze brain signals, and the stability of performance over time (Murphy et al., 2016).

A BMI is a system that translates cerebral activity into commands to communicate with an external device, bypassing the normal neuromuscular pathways (Wolpaw et al., 2002; Aydin et al., 2018). There are various techniques to register brain signals, but the non-invasive neuroimage modality most widely used in BMI applications is electroencephalography (EEG) because of its high temporal resolution, low cost, and mobility (Flores et al., 2018; Xiao et al., 2019). Among EEG-based BMIs, the P300 paradigm is one of the most popular techniques for building applications with multiple options because it allows achieving high accuracies without the need for long calibration sessions (Hwang et al., 2013; De Venuto et al., 2018). Compared with other paradigms, P300-based BMIs have higher bit rates than motor imagery interfaces, while the stimulation technique for evoking P300 potentials is less visually fatiguing than the method used to elicit steady-state visually evoked potentials (Cattan et al., 2019).

The P300 signal is an event-related potential (ERP) component observed in the electroencephalogram elicited about 300 ms after the perception of an oddball or relevant auditory, visual, or somatosensory stimulus (Cattan et al., 2019). Typically, in a P300-based BMI, characters, syllables, or icons presented on a computer screen flash randomly one at a time while the user focuses attention on one particular graphical element (target stimulus). Each flashing stimulus represents an option, action, or command that the system can execute. The user selects one option of the interface by counting or performing a cognitive task every time the target stimulus is highlighted. Because the target option flashes randomly, this stimulus produces a P300 evoked potential synchronized with the flickering event in the timeline. In this way, a P300-based BMI identifies which option is evoking an ERP to decode the user's intentions and perform the desired action.

Numerous published works have reported examples of P300-based BMIs for communication and control, including spellers (Kleih et al., 2016; Poletti et al., 2016; Okahara et al., 2017; Flores et al., 2018; Guy et al., 2018; Deligani et al., 2019; Shahriari et al., 2019), authentication systems (Yu et al., 2016; Gondesen et al., 2019), assistive robots (Arrichiello et al., 2017), smart home environments (Achanccaray et al., 2017; Masud et al., 2017; Aydin et al., 2018), neurogames (Venuto et al., 2016), remote vehicles (De Venuto et al., 2017; Nurseitov et al., 2017), wheelchairs (De Venuto et al., 2018), and robotic arms (Tang et al., 2017; Garakani et al., 2019). Because the development of assistive technologies for motor-impaired people is one of the major purposes of BMI research, some groups have evaluated similar applications in clinical environments on people with neurological disorders or reduced mobility. Regarding medical applications, we can find P300-based BMIs for ALS (Liberati et al., 2015; Schettini et al., 2015; Poletti et al., 2016; Guy et al., 2018; Deligani et al., 2019; Shahriari et al., 2019; McFarland, 2020), Alzheimer's (Venuto et al., 2016), spinocerebellar ataxia (Okahara et al., 2017), and post-stroke paralysis (Kleih et al., 2016; Achanccaray et al., 2017; Flores et al., 2018). Recently, P300-based BMIs have also been proposed for rehabilitation contexts (Kleih et al., 2016), and diagnosis/evaluation purposes (Poletti et al., 2016; Venuto et al., 2016; Deligani et al., 2019; Shahriari et al., 2019).

Some studies have stated the benefits of orthoses for ALS patients (Tanaka et al., 2013; Ivy et al., 2014); however, the implementation of BMI-controlled robotic hand-orthoses for this target population remains underexplored in comparison to the application of these systems for other neuromotor disorders. Moreover, most of the recent published BMIs for ALS are designed for communication purposes (Vaughan, 2020). Similarly, while the employment of BMI-controlled hand-orthoses is well-known in other neuromotor conditions (e.g., stroke recovery), the effect of the use of these systems in ALS patients remains poorly investigated. A critical step toward the development of practical robotic neuroprostheses for people with ALS is the evaluation of this technology in different scenarios. It is essential to determine if ALS patients can operate this particular mind-controlled application and evaluate the possible effect of a hand-orthosis on the user's experience and performance.

This work presents the development and evaluation of a P300-based BMI coupled with a robotic hand-orthosis device. The purpose of this system is to assist people with ALS to perform movements of individual fingers of one hand, or more complex tasks that involve a sequence of actions of one or more fingers. Eighteen healthy participants and eight ALS patients conducted an experiment in which they tested the proposed BMI selecting a sequence of actions that the robotic hand-orthosis executed. In the evaluation of this BMI, we considered six types of operations: the flexion-extension of individual fingers, and the flexion-extension of the five fingers simultaneously.

In the experiments, we recorded the data used in the training phase of the BMI, and the EEG signals measured during the online tests. The training data was used to evaluate offline the performance of the classification model implemented in the BMI to discriminate between target and non-target epochs. Additionally, we analyzed the P300 responses of the participants to determine if there are subjects without clear evoked potentials. In the online tests, we calculated the classification accuracy and the selection times of the BMI. It is important to say that some selections were made without connecting the hand-orthosis to the system to evaluate the effect of the robotic device in the online accuracy of the BMI.

To our knowledge, our system is the first P300-based BMI that allows ALS patients to perform sequences of movements of individual or two or more digits simultaneously; it is important to consider the advantage of our system to allow the individual movement of the digits since ALS is associated with the degeneration of the corticospinal tract (Sarica et al., 2017) that allows to perform the fine finger motor tasks (Levine et al., 2012). Besides, being a P300-based system, the calibration precises a

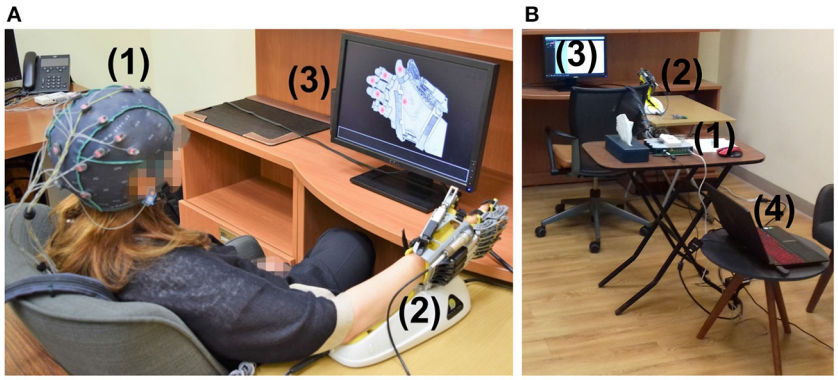


FIGURE 1 | Hardware components of the BMI: (1) EEG recording system (EEG cap, active electrodes, and amplifier), (2) Hand of Hope orthosis, (3) monitor to display the GUI, and (4) computer to process EEG signals, synchronize the stimuli, and send control commands. **(A)** System elements in direct contact with users. **(B)** Setup of the BMI elements.

minimum time consuming calibration, reducing the fatigue of patients in comparison with other systems.

Another contribution of this work is the dataset obtained in the experimental sessions of the proposed BMI. This dataset contains the training data and the online recordings of 26 participants. The calibration samples are useful to evaluate different machine learning models of P300-based BMIs, whereas the online signals can be used to test practical systems without the need for real participants. The relevance of this database resides in the importance of providing high-quality EEG observations that represent both control and ALS groups. Any researcher may evaluate other P300-based BMIs and verify if their proposals can correctly identify the user's intentions in online conditions.

The remainder of this paper is divided into three sections. Section 2 describes the hardware and software components of the mind-controlled hand-orthosis, and the experimental setup under which we tested the BMI. Section 3 shows the results obtained from the system evaluation, while section 4 discusses the implications of the results and the conclusions derived from this work.

2. MATERIALS AND METHODS

2.1. Brain-Machine Interface

The proposed system consists of a P300-based BMI coupled with a Hand Of Hope robotic arm (Rehab-Robotics Company, China). This hand-orthosis is a therapeutic device with five DC linear motors designed initially for the rehabilitation of post-stroke patients (Aggogeri et al., 2019). There is a detailed description of the Hand of Hope and its functionality in Ho et al. (2011). To communicate the orthosis with the BMI, we enabled a wireless communication channel to send the position of each motor during the execution of one movement or sequence of movements. In this way, the user selects one action to perform with the hand-orthosis using the P300-based interface.

Figure 1 sketches the components of the mind-controlled hand-orthosis, and how the users interact with them. The main hardware components of the interface are:

- An EEG recording system (a g.GAMMASYS active wet electrode arrangement and a g.USBamp amplifier provided by g.tec medical engineering GmbH, Austria). For this study, the sampling rate was 256 Hz, and we used eight monopolar electrodes, placed according to the 10–20 international system at positions Fz, Cz, P3, Pz, P4, PO7, PO8, and Oz. The ground electrode was located at AFz, and the reference electrode on the right earlobe.
- A Hand of Hope robotic arm. The users can wear the robotic device on any hand.
- A monitor that displays the graphical user interface (GUI) of the BMI.
- A computer that processes the EEG signals, synchronizes the stimulus presentation, and sends the control commands to the hand-orthosis.

The software elements of this system, including the GUI, were implemented in-house using C++.

The GUI of the BMI (shown in **Figure 2**) provides the instructions to operate the system, presents the flashing elements, and displays visual feedback. In this GUI, gray circles positioned on a graphical illustration of the hand-orthosis represent the available options (i.e., actions or movements of the robotic device). Since the orthosis can be used on any hand, the GUI can display the image of a left or right hand, according to the side where the robotic device would be placed.

It is possible to program different movements or actions for the hand-orthosis. The system can move each finger independently or perform multiple movements at the same time. For this study, we evaluated the BMI using six options: the individual flexion-extension of each finger, and the simultaneous flexion-extension of the five fingers. The five gray circles placed over the fingers represent the individual movements, whereas the circle over the palm corresponds to the hand opening and closing.

In this system, the stimulation method used to elicit evoked responses is the dummy face pattern (Chen et al., 2015, 2016), which consists of a yellow happy face icon that replaces for a short time a gray circle selected randomly. In one flashing cycle, the happy face icon is shown for 75 ms, and then all

the gray circles appear for 75 ms (see **Figure 3**). The users are instructed to choose freely one movement or action of the robotic device by counting how many times the happy face is displayed on the desired option. If the system detects a P300 response for one action, the flashing stops for 4 s, while the hand-orthosis performs the corresponding movement. Then, the interface restarts the random flashing and waits for another evoked response. The same routine is repeated continuously during the regular operation of the BMI.

The detection of evoked responses consists of a sequence of processing steps necessary to extract relevant information from the measured EEG signals. **Figure 4** summarizes the different stages implemented in our BMI to analyze and classify electrophysiological data. Firstly, when one option flashes, the system extracts the EEG epoch (or trial) around this event and applies some pre-processing and feature extraction techniques on this data segment. Then, a classification model evaluates the computed characteristics to obtain the label that represents the class of the processed epoch (target or non-target). A third class is also considered in this design (artifact) to indicate if a trial is contaminated by noise or muscle artifacts. Finally, the BMI processes this label to determine whether the flashing stimulus is eliciting evoked responses. If there is a P300 evoked potential, the BMI sends the respective control signals to the robotic device. This processing pipeline is based on the classification approach described in Mendoza-Montoya (2017).

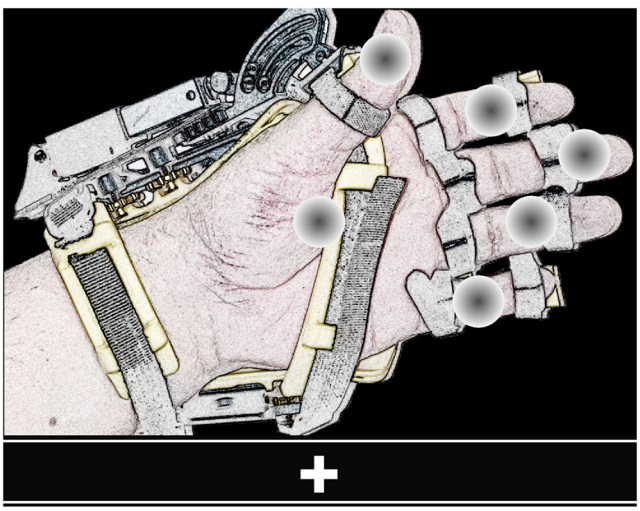


FIGURE 2 | View of the GUI. The screen shows six flashing gray circles (possible options) placed over the image of a left or right hand wearing the orthosis. The bar located at the lower part of the GUI presents the instructions to operate the BMI and feedback.

The following describes the processing stages of the BMI, and the component of the interface that interacts with the robotic device. Also, we present the calibration routine implemented to train the system.

2.1.1. EEG Pre-processing

In the pre-processing stage, the BMI extracts three band-limited components using FIR filters with cut-off frequencies between 4 and 14, 20–40, and 4–40 Hz. Then, when one flashing occurs, the samples around the time-window of the event are separated so that the epoched signals contain 800 ms of post-stimulus samples, starting from the stimulus onset. The result of this processing step are signals $X^{4-14} = \left[x_e^{4-14}(t)\right] \in \mathbb{R}^{n_e \times n_t}$, $X^{20-40} = \left[x_e^{20-40}(t)\right] \in \mathbb{R}^{n_e \times n_t}$, and $X^{4-40} = \left[x_e^{4-40}(t)\right] \in \mathbb{R}^{n_e \times n_t}$, where e represents the electrode position ($e = 1, 2, 3, \ldots, n_e$), n_e is the number of electrodes or channels, t is the time index ($t = 1, 2, 3, \ldots, n_t$), and n_t is the number of samples.

The next step is the epoch validation, which is necessary to determine if the EEG trial is not contaminated by muscle artifacts or other sources of noise. Here, the BMI calculates the peak-to-peak voltage v_e^{pp}, the standard deviation σ_e, and the power ratio r_e of each channel as follows:

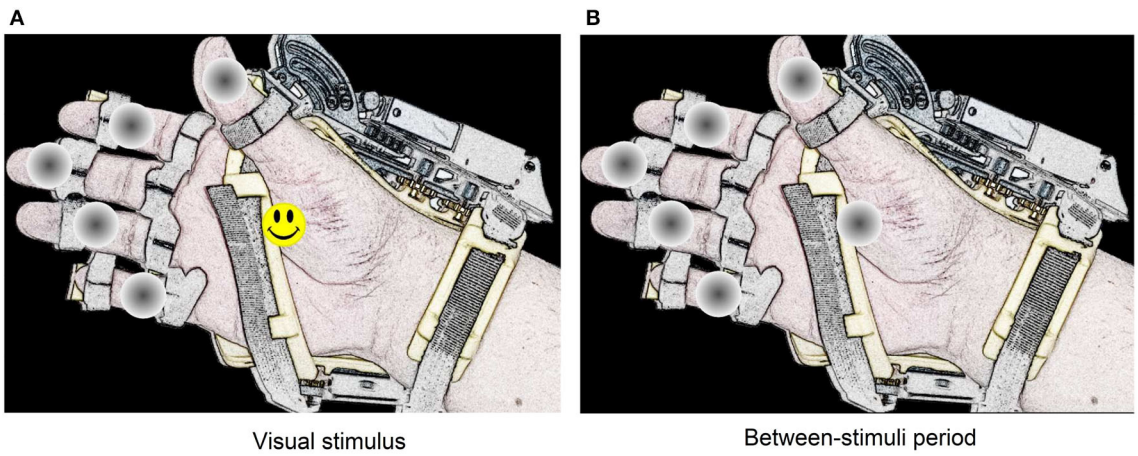


FIGURE 3 | Representation of the dummy face pattern method for visual stimulation. The visual stimulus remain active during 75 ms on one option selected randomly **(A)**. Between each stimulus, there is a period of 75 ms where all circles remain gray colored **(B)**.

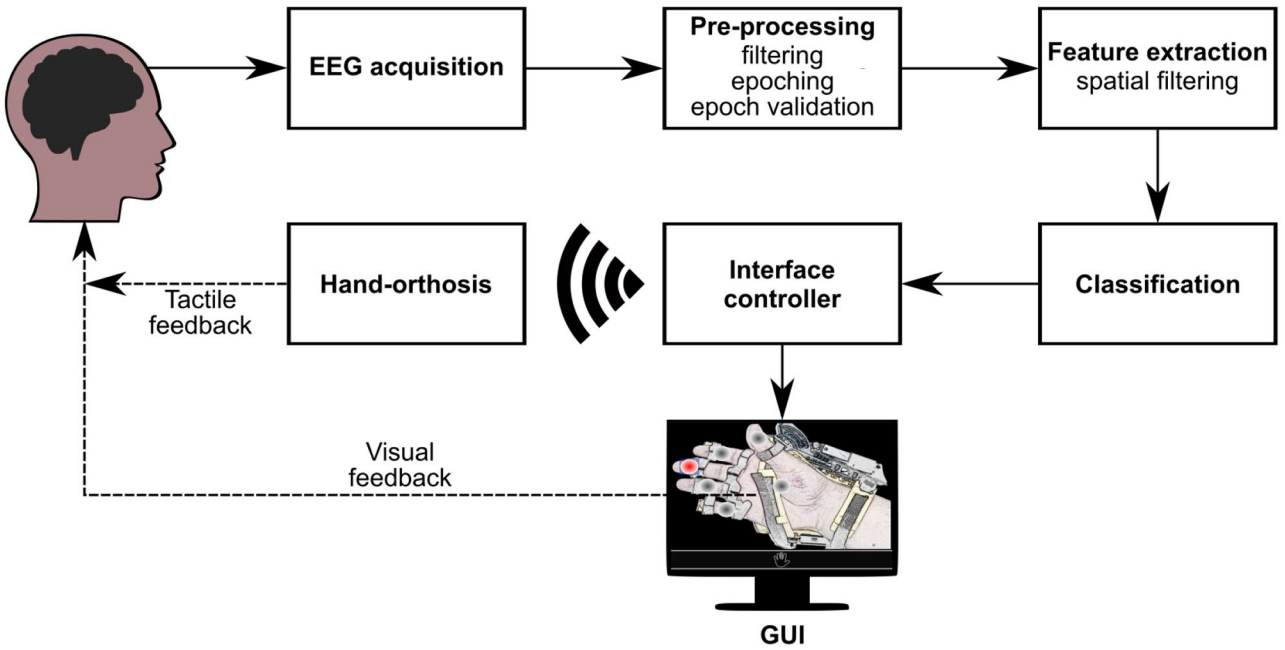


FIGURE 4 | Processing stages of the BMI. Once acquired the EEG signals, they go through the pre-processing, feature extraction, and classification stages. Based on the results of the classifier, the interface controller synchronizes the visual stimuli and sends control commands via WiFi to the orthosis. If the BMI detects a target option, the GUI provides visual feedback to the user, and the orthosis provides tactile feedback.

$$v_e^{pp} = \max_t \left(x_e^{4-40}(t)\right) - \min_t \left(x_e^{4-40}(t)\right), \quad (1)$$

$$\sigma_e = \sqrt{\frac{1}{n_t - 1}\sum_{t=1}^{n_t}\left(x_e^{4-40}(t) - \mu_e\right)^2}, \quad (2)$$

$$r_e = \frac{\sum_{t=1}^{n_t}\left(x_e^{20-40}(t)\right)^2}{\sum_{t=1}^{n_t}\left(x_e^{4-40}(t)\right)^2}, \quad (3)$$

where

$$\mu_e = \frac{1}{n_t}\sum_{t=1}^{n_t} x_e^{4-40}(t). \quad (4)$$

The system classifies as "artifact" any epoch with one or more channels for which $v_e^{pp} \geqslant 200\ \mu$V, $\sigma_e \geqslant 50\ \mu$V, or $r_e \geqslant 0.7$. In this case, the trial is not processed and evaluated by the machine learning model of the BMI. On the other hand, if the epoch passes the validation, i.e., the calculated metrics for all channels are below the threshold levels, the system downsamples X^{4-14} using a decimation factor of four to obtain signal $Y = \left[y_e(t)\right] \in \mathbb{R}^{n_e \times \hat{n}_t}$, where $\hat{n}_t$ is the number of time points after the downsampling.

2.1.2. Feature Extraction

The system implements an algorithm of spatial filtering based on canonical correlation analysis (CCA) for feature extraction. This approach is effective in reducing the data dimensionality and increasing the classification accuracy (Spüler et al., 2014; Mendoza-Montoya, 2017). Spatial filtering is a technique that finds linear combinations or projections of a set of signals in such a way that the new signals in the projected space have better separability between classes or another improved property. Given column vector $w = [w_e] \in \mathbb{R}^{n_e}$ with n_e spatial weights, the projected signal $\tilde{y}(t)$ is obtained as follows:

$$\tilde{y}(t) = \sum_{e=1}^{n_e} w_e y_e(t). \quad (5)$$

In our BMI, the spatial weights increases the correlation between epochs of the target option and the expected ERP response of this class. Let $Y^{\text{target}} = \left\{Y_1, Y_2, Y_3, \ldots, Y_{n_{\text{target}}}\right\}$ be a set with n_{target} pre-processed observations free of artifacts of the target class obtained from raw calibration data ($Y_k = \left[y_{e,k}(t)\right] \in \mathbb{R}^{n_e \times \hat{n}_t}$). The average ERP waveform $\overline{Y}^{\text{target}} \in \mathbb{R}^{n_e \times \hat{n}_t}$ of these observations is:

$$\overline{y}_e^{\text{target}}(t) = \frac{1}{n_{\text{target}}}\sum_{k=1}^{n_{\text{target}}} y_{e,k}(t). \quad (6)$$

The training epochs of the target class and their average ERP waveform are concatenated to build matrices $U = [Y_1, Y_2, \ldots, Y_{n_{target}}]^T$ and $V = [\overline{Y}^{\text{target}}, \overline{Y}^{\text{target}}, \ldots, \overline{Y}^{\text{target}}]^T$ of dimensions $\left(\hat{n}_t \cdot n_{\text{target}}\right) \times n_e$, where T denotes transpose. CCA is applied to calculate vectors w and $\widetilde{w}$ that maximize the correlation between Uw and $V\widetilde{w}$. Here, the system uses w as spatial filter to transform the pre-processed epoch.

Because CCA produces n_e spatial filters, the system selects the best n_w projections which correspond to the highest correlation

values ($1 \leqslant n_w \leqslant n_e$). To evaluate the system performance, we set $n_w = 4$. In this way, after the spatial filtering, the BMI obtains the projected signal $\tilde{Y} \in \mathbb{R}^{n_w \times \hat{n}_t}$.

2.1.3. Classification

To classify one flashing event, the machine learning model of the BMI evaluates the corresponding signal $\tilde{Y}$ and returns a label or category $L \in \{\text{target}, \text{non-target}\}$, indicating whether the flickering option is a target stimulus. This operation is only applied to trials free of noise or artifacts. For non-valid epochs, the assigned label is "artifact."

The system uses the regularized version of the linear discriminant analysis (RLDA) (Lotte and Guan, 2009) to distinguish between the target and non-target epochs. This binary model has been employed previously to detect P300 potentials (Zhumadilova et al., 2017) and classify other electrophysiological responses (Cho et al., 2018). In this stage, the classifier evaluates only a small subset $Z = \{z_1, z_2, \ldots, z_{n_f}\}$ of n_f features, selected from the $n_w \times \hat{n}_t$ spatially filtered variables ($z_i \in \{\tilde{y}_e(t)\}$). This dimensionality reduction is necessary to prevent overfitting and reduce the complexity of the machine learning model (Tyagi and Nehra, 2017).

During the system calibration, the BMI chooses the characteristics to evaluate in the classification stage using the forward-backward stepwise (SW) method for variable selection (James et al., 2015). This algorithm starts with an empty classification model without variables and incorporates the one that contributes best to the model performance according to a scoring criterion. Then, the best feature that is not in the model and improves the performance criterion significantly is incorporated. If none of the candidate variables help to enhance the classifier, the model is not modified. In the next step, the variable that is in the model that may be excluded without reducing the actual scoring significantly is removed. Again, the feature set is not altered when it is not possible to discard one feature without worsening the model. These steps are repeated until no more changes in the feature set are possible. In this framework, the features selection and the model training are performed simultaneously.

2.1.4. Interface Controller

The label obtained in the classification stage might be used directly to determine the action or movement to produce with the robotic device. However, because the accuracy of the machine learning model is typically below 90%, the risk of executing an incorrect action is high. It is essential to consider that there is only one target stimulus and multiple non-target options so that before evaluating an epoch of the desired option, the model must detect correctly multiple instances of the non-target class. For this reason, the system processes the history of labels to determine if there is enough evidence that the user wants to select one particular option.

The interface controller is the element of the BMI that receives the labels of the flashing events and determines which action must perform the hand-orthosis. Additionally, it generates the control signals necessary to perform the selected movements or actions and synchronizes the state of the GUI to produce visual feedback. This component decides when the hand-orthosis must be activated and which movement or sequence of movements it must execute.

When the system processes one flashing event, the interface controller evaluates the number of times that each option has been classified as target and non-target responses. Only the last ten flashing events of each flashing symbol free of artifacts are considered in this counting. One option is selected if the following conditions are satisfied:

- The corresponding gray circle of the analyzed option has flashed at least five times (minimum number of processed epochs).
- At least 70% of the flashing events of this option has been classified as target stimuli (target class threshold).
- The responses of each of the other flashing circles have been classified 60% of the time as non-target (non-target class threshold).

If the controller detects a P300 response for one particular option, the flashing sequence is interrupted, providing visual feedback to the user about the selection. Subsequently, the hand-orthosis executes the chosen routine, and the flashing sequence restarts for another selection.

2.1.5. Calibration Routine

The operation of the BMI requires a set of spatial filters and a classification model to process and evaluate the epochs of the flashing events. To find these components, the system provides a calibration routine in which the user focuses attention on target options while the BMI records the subject's brain signals. This routine replicates the operational conditions of the BMI without activating the hand-orthosis. It uses the same GUI with six options, the stimulation method is the dummy face pattern, and the happy face icon appears for 75 ms alternating with 75 ms of no visual stimulus. Because the hand-orthosis is not necessary to train the interface, this device is disabled, and the user is not instructed to wear it.

The calibration routine is divided into eight training sequences or runs. A run (shown in **Figure 5**) starts with a fixation cross to indicate that a training sequence has begun. Then, the interface presents a target option (selected randomly by the interface), followed by short preparation time. Next, the options flash randomly one after another for 30 s. Here, the user must count mentally how many times the happy smile icon appears on the specified target option. Finally, the user rests for a few seconds before the next run. At the end, the training dataset contains 264 epochs of the target class and 1,320 trials of the non-target class.

After completing the calibration routine, the system processes and validates the dataset to train the machine learning model of the BMI (see **Figure 6**). Firstly, the system pre-processes the complete dataset to obtain downsampled epochs free of artifacts of both classes. Next, the spatial filters are calculated using the set of observations of the target class. Then, the calculated filters are applied to the extracted epochs of both classes. Finally, the spatially filtered observations are used to find the optimal subset

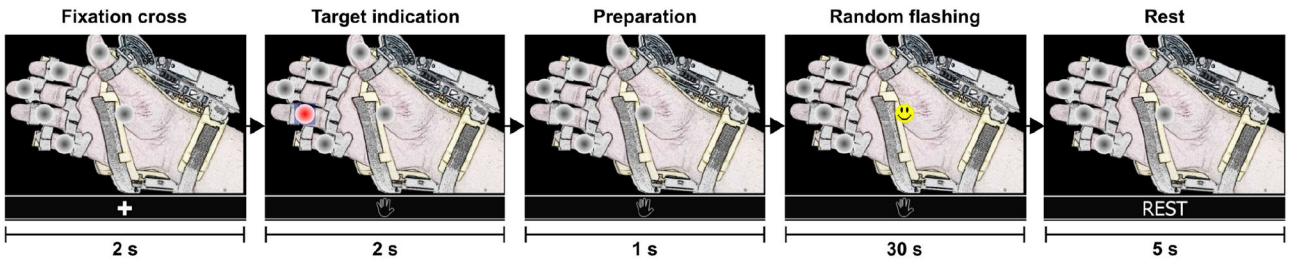


FIGURE 5 | Steps of a single training sequence (training run). The complete calibration consists of eight runs and lasts 320 s. At the end of the calibration routine, the training dataset contains 264 epochs of the target class and 1,320 trials of the non-target class.

FIGURE 6 | Processing stages for system calibration. The information contained in a calibration dataset is pre-processed and analyzed to obtain a set of spatial filters and a classification model. These two components are necessary to operate the BMI and control the hand-orthosis.

of features of the classifier and the parameters of this model. Once the classification model is trained, the BMI is ready to operate online and send control commands to the hand-orthosis.

2.2. Participants

To evaluate the proposed mind-controlled hand-orthosis, we conducted an experiment in which HS and people with ALS tested the brain-machine interface. In this study, we included eighteen healthy participants (10 females and eight males, aged between 19 and 63 years old, mean age 32.7) and eight ALS patients (three females and five males, aged between 49 and 72 years old, mean age 59.6), all with normal or corrected vision. **Table 1** shows the age range of each participant, and the characteristics of the ALS patients. Study subjects had no previous experience with any brain-machine interface.

ALS participants were selected from the patients attending the TecSalud ALS Multidisciplinary Clinic (Martínez et al., 2020) considering the disease duration and disability level as inclusion criteria. According to this criterion, the eight participants had, at the time of the tests, a disease duration from 2 to 3 years, and a general disability level ranged from mild to moderate (according to the ALSFRS-R scores). Both, HS and ALS groups volunteered for the study and provided informed consent before the experimental sessions. This study followed the ethical principles of the World Medical Association (WMA) Declaration of Helsinki (WMA, 2013).

2.3. Experimental Design

The experiments were carried out in a dedicated medical room at Zambrano-Hellion Medical Center. HS and patients who could walk without help or a wheelchair were asked to sit in a comfortable chair approximately one meter apart from the 22 inch LCD monitor of the BMI. For patients that needed assistance, the room space was adapted to accommodate a wheelchair close to the robotic device in front of the monitor. Before starting the experiments, participants were informed about the general instructions of the different tasks and were asked to avoid unnecessary movements when they had to focus attention on the interface.

Figure 7 summarizes the different stages of one experimental session. After placing the EEG cap and preparing the wet electrodes, the experimenter instructed the participants to calibrate the BMI and perform a short free validation. In this test, subjects selected freely any option of the interface and notified if the system detected the desired action correctly. The purpose of the free selections was to obtain information about the detection times and demonstrate the users that the BMI is effectively responding to their intentions. Participants repeated at least three times the free target selection before continuing with the experiment.

In the next stages of the experiment (online tests), subjects were indicated to focus attention on the specified target option until the BMI recognized a P300 response for one of the flashing elements. All online attempts (shown in **Figure 8**) are similar to the calibration runs. The interface presented a fixation cross to indicate the beginning of a test run, followed by the presentation of the target option and preparation time. Then, the random flashing started, and the BMI tried to recognize an evoked response. If the system detected in <30 s the correct option, the hand-orthosis performed the selected movement; otherwise, nothing happened. Finally, there were 5 s of resting time before starting another attempt.

Some online runs were performed with the robotic device disabled. In these cases, subjects did not wear the hand-orthosis, and the system did not send the control signals to the device. **Table 1** indicates the number of online attempts performed by each participant with and without the Hand of Hope. In this way, we collected three datasets for each participant, the calibration data, the online test data without the robotic device, and the EEG recordings with the hand-orthosis.

TABLE 1 | Characteristics of the participants.

Subject	Age range (years)	Target detection attempts		
		Without orthosis	With orthosis	Total
HS1	21-25	12	18	30
HS2	61-65	18	18	36
HS3	21-25	12	18	30
HS4	16-20	12	12	24
HS5	16-20	12	12	24
HS6	51-55	18	18	36
HS7	46-50	12	12	24
HS8	51-55	18	18	36
HS9	26-30	18	12	30
HS10	26-30	12	12	24
HS11	21-25	18	18	36
HS12	61-65	18	18	36
HS13	16-20	18	6	24
HS14	16-20	18	18	36
HS15	21-25	18	18	36
HS16	26-30	18	12	30
HS17	21-25	18	18	36
HS18	21-25	18	6	24

					ALSFRS-R	Years from symptoms onset	Hand motor impairment
ALS1	46-50	0	12	12	44	2	mild
ALS2	56-60	0	12	12	35	2	moderate
ALS3	61-65	0	6	6	40	2	moderate
ALS4	56-60	0	12	12	33	2	advanced
ALS5	46-50	18	12	30	26	2	advanced
ALS6	71-75	18	18	36	35	2	moderate
ALS7	61-65	18	18	36	42	2	moderate
ALS8	61-65	12	12	24	39	3	moderate

This table specifies the age range of each participant and the number of online validation runs performed for each condition of orthosis usage. The column "Total" indicates the total number of validation runs performed among both conditions. For patients, the last three columns indicate the ALS Functional Rating Scale Score (ALSFRS-R) score, the years from the onset of the ALS symptoms, and the level of motor impairment of the hands measured as mild (no observable to sporadic symptoms), moderate (visible symptoms), or advanced (no residual movement).

FIGURE 7 | Different stages of the experiment designed for evaluating the mind-controlled hand-orthosis. An experimental session started with the subject preparation and system setup. Then, the participant trained the BMI and tested the interface freely. Finally, the experiment concluded with the online tests. A complete experiment lasted approximately from 30 to 55 min.

2.4. Data Analysis

2.4.1. ERP Analysis

The calibration data recorded in our experiments was used to analyze the ERP responses of each participant. In this study, we pre-processed and validated the training epochs of the target and non-target classes to calculate the average waveforms of both conditions. To determine the ERPs, we used the filtered signals obtained with the bandpass filter of 4–40 Hz. We considered 200 ms of pre-stimulus samples and 800 ms of post-stimulus time points.

Significant ERP peaks were identified through a statistical test of the ERP amplitude at each time point and channel with the

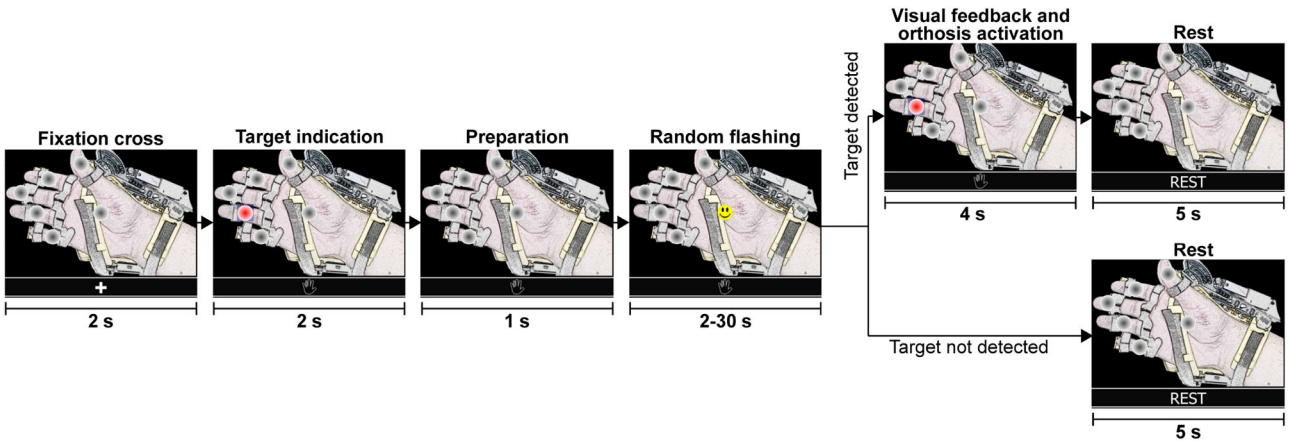


FIGURE 8 | Graphical representation of a single test run. During the random flashing, the BMI tried to detect an evoked potential while the participant focused attention on the specified target option. If the system detected the correct option, the robotic device performed the selected action. If the target was not detected within 30 s during the random flashing, the system went directly to the rest period before starting a new test run with the fixation cross.

corresponding probability density function (PDF) of the pre-stimulus interval. We estimated the PDF of the pre-stimulus segment of each channel with the non-parametric kernel density estimation method (Bowman and Azzalini, 1997). The upper and lower limits of the PDFs were then computed for a significance level of $\alpha = 0.05$, i.e., significant ERP responses are those for which the probability values under the PDF of the corresponding pre-stimulus are higher than $1 - \alpha/2$ or smaller than $\alpha/2$.

Significant ERP responses in the target class indicate that the interface is eliciting evoked potentials when the subject perceives a flashing event of a target option. On the other hand, it is expected not to observe significant evoked potentials in the non-target class because the subject is not attending these events.

2.4.2. Classification Model Evaluation

In this study, we evaluated the accuracy of the machine learning model of the BMI for each subject by applying five-fold cross-validation on the calibration data (Berrar, 2019). This method is useful to estimate the prediction error and accuracy of a model when the number of available observations is limited, and it is not possible to split the complete dataset into training data and test data. For this assessment, we report the accuracy acc_i of each class $i \in \{\text{target}, \text{non-target}\}$ (the proportion of samples of class i predicted in this class correctly), and the weighted model accuracy $acc_w = 0.5 \times (acc_{\text{target}} + acc_{\text{non-target}})$. We used the weighted accuracy because the training data sets are highly unbalanced, and we want to avoid a bias toward the non-target class.

Additionally, the significance levels of the model accuracies were calculated with a permutation test (Good, 2006). In this methodology, the null hypothesis indicates that observations of both classes are exchangeable so that any random permutation of the class labels produces similar accuracies to the obtained with the non-permuted data. The alternative hypothesis is accepted when the model accuracy is an extreme value in the empirical distribution built with m random permutations. When the alternative hypothesis is accepted, we can say that the cross-validated accuracy is above the chance level.

2.4.3. Online Evaluation

We assessed the online BMI performance in terms of selection accuracies, detection times and ITR. These parameters are computed through Equations (7)–(9), where acc_{online} is the online accuracy, n_{sel} is the number of correctly selected targets, n_{att} is the number of attempts to select a target or test runs, B is the information-rate transmitted (bits), n_c is the number of flashing circles, and t_{avg} is the average time from target indication to target selection (detection time in seconds).

$$acc = \frac{n_{sel}}{n_{att}} \times 100\%. \tag{7}$$

$$B = \log_2 n_c + (acc)(\log_2 acc) + (1 - acc)\log_2 \frac{1 - acc}{n_c - 1}. \tag{8}$$

$$ITR = 60 \times \frac{B}{t_{avg}}. \tag{9}$$

3. RESULTS

3.1. ERP Responses

Figure 9 shows the results of the ERP analysis for one of the healthy subjects (HS6) and one of the ALS patients (ALS2). This analysis is presented for all channels separately for the target and non-target conditions. For the two participants, significant positive and negative peaks ($p < 0.05$, two-tail test) are observed in the ERP for the target condition (top figures), while no significant ERP peaks ($p > 0.05$, two-tail test) are observed in the non-target condition (bottom figures).

For the healthy subject HS2, the ERP in the target condition shows (i) a positive peak between 250 and 450 ms in all channels (the P300 response), (ii) a negative peak between 450 and 550 ms in the frontal Fz and the central Cz channels (possibly a late negativity), and (iii) an early negativity around 200 and 250

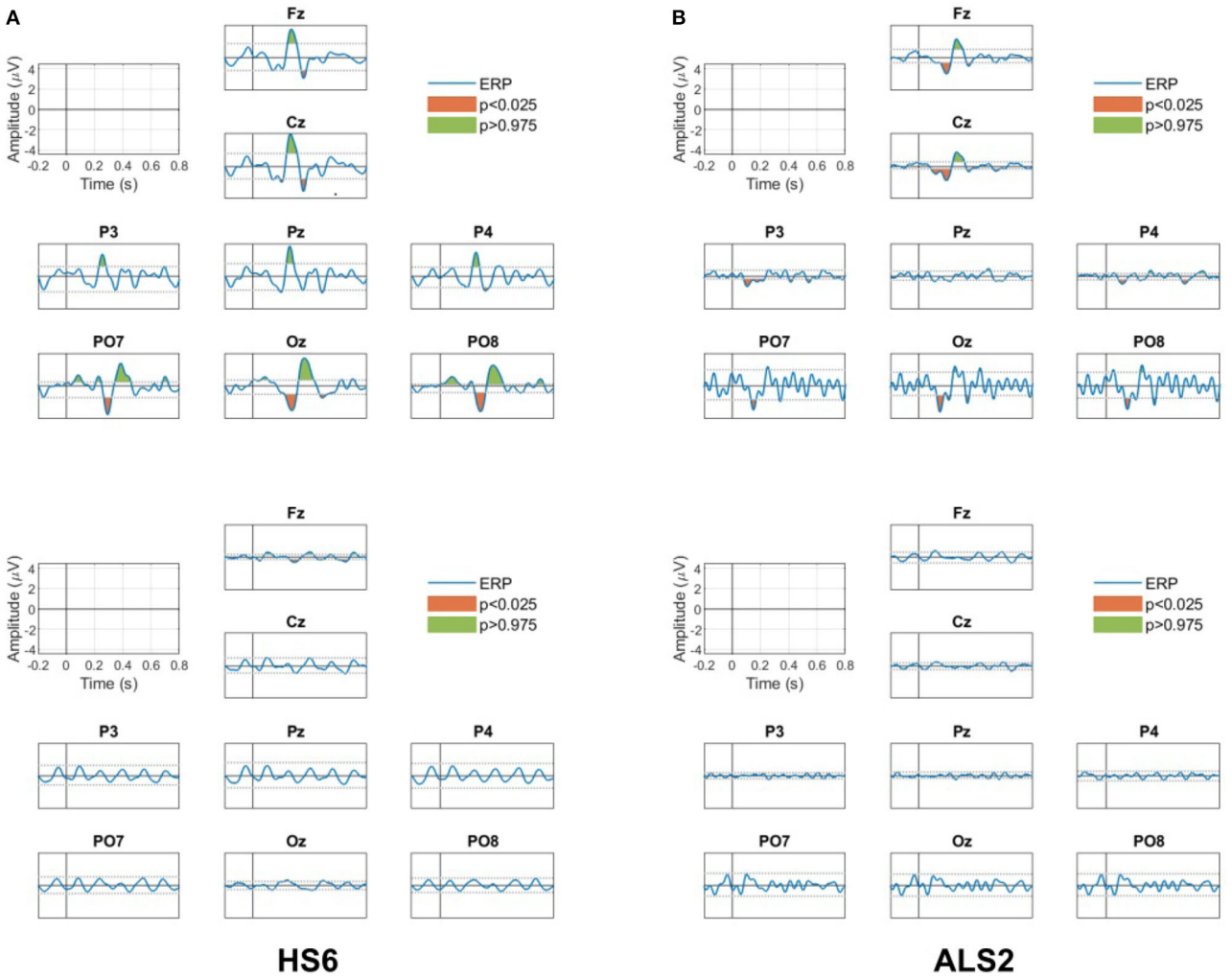


FIGURE 9 | ERP responses for all channels for the target (upper panels) and non-target (bottom panels) conditions for **(A)** the healthy participant HS6 and **(B)** the patient ALS2. Green and red areas in the ERP for the target condition are the positive and negative peaks that presented significant differences ($p < 0.05$, two tail test) with the estimated PDF of the baseline period. No significant peaks are observed in the ERP for the non-target condition.

ms in the parieto-occipital (PO7 and PO8) and occipital (Oz) channels. Note that none of these features are observed in the non-target condition.

For the patient ALS2, the ERP in the target condition shows (i) the positive peak representing the P300 response between 250 and 450 ms in the frontal Fz and the central Cz channels, and (ii) an early negativity around 200 and 250 ms in the frontal and the central (Fz and Cz), the parieto-occipital (PO7 and PO8) and the occipital (Oz) channels. Note that these significant peaks are not observed in the non-target condition.

Similar observations are also present in the rest of the participants and indicate the existence of significant task-related evoked activity that is used by the proposed BMI system to recognize the stimulus the user is attending.

3.2. Classification Model Accuracy

Table 2 contains the classification accuracies estimated with five-fold cross-validation for each participant. The mean accuracy for the target class was 78.7%, for the non-target class was 85.7%, and the weighted accuracy was 82.2%. Only the model performance for two participants was below 70% (HS17 and ALS3), whereas three participants obtained accuracies above 90% (HS6, HS7, and HS10). The best classifier performance was 95.8%, and the worst was 66.5%. All these results are similar to those reported in other similar works (Wang and Chakraborty, 2017; Won et al., 2018).

In the permutation tests, the classification accuracies for all participants were significant ($p < 0.001$, 10,000 random permutations). These results indicate that the machine learning model implemented in our BMI can discriminate between EEG epochs of the target and non-target classes. However, if we want to avoid selection errors in the online operation, it is important to consider a multi-trial strategy because the error rates are not zero. For this reason, the interface controller processes consecutive labels returned by the classification stage to determine the desired option.

TABLE 2 | Classification accuracies (%) estimated with cross-validation for the target and non-target classes.

Subject	Target	Non-target	Mean
HS1	76.1	85.1	80.6
HS2	76.2	85.3	80.8
HS3	85.3	90.4	87.8
HS4	81.4	88.6	85.0
HS5	78.7	85.3	82.0
HS6	93.3	98.4	95.8
HS7	90.5	95.2	92.8
HS8	86.3	90.5	88.4
HS9	83.8	87.9	85.9
HS10	87.0	94.4	90.7
HS11	77.2	85.5	81.4
HS12	75.7	81.3	78.5
HS13	72.5	78.9	75.7
HS14	87.2	91.2	89.2
HS15	71.9	79.4	75.7
HS16	73.7	80.4	77.1
HS17	61.6	76.1	68.9
HS18	71.2	81.5	76.3
ALS1	80.9	87.6	84.2
ALS2	71.5	76.3	73.9
ALS3	63.2	69.8	66.5
ALS4	79.5	86.7	83.1
ALS5	73.0	83.4	78.2
ALS6	86.5	91.5	89.0
ALS7	84.5	89.9	87.2
ALS8	77.8	86.5	82.2
Mean	78.7	85.7	82.2
Std	7.8	6.5	7.1

The fourth column indicates the model accuracy (mean value).

TABLE 3 | Online classification performance obtained in the evaluation of the mind-controlled hand-orthosis.

	Accuracy (%)		
ID	Without orthosis	With orthosis	Total
HS1	75.00	83.33	80.00
HS2	100.00	88.89	94.44
HS3	100.00	100.00	100.00
HS4	100.00	100.00	100.00
HS5	100.00	100.00	100.00
HS6	100.00	100.00	100.00
HS7	100.00	100.00	100.00
HS8	100.00	100.00	100.00
HS9	100.00	91.67	96.67
HS10	100.00	100.00	100.00
HS11	88.89	100.00	94.44
HS12	94.44	88.89	91.67
HS13	83.33	83.33	83.33
HS14	100.00	100.00	100.00
HS15	72.22	77.78	75.00
HS16	66.67	58.33	63.33
HS17	50.00	44.44	47.22
HS18	88.89	66.67	83.33
ALS1	NA	83.33	83.33
ALS2	NA	75.00	75.00
ALS3	NA	100.00	100.00
ALS4	NA	100.00	100.00
ALS5	77.78	66.67	73.33
ALS6	100.00	100.00	100.00
ALS7	100.00	88.89	94.44
ALS8	100.00	100.00	100.00
Mean	90.78	88.35	89.83
Std	14.12	15.38	13.87

Results are reported separately for each condition of orthosis usage (with or without orthosis). The fourth column (total) includes the results for all the test runs regardless of if the orthosis was used or not. NA indicates absence of validation runs under that experimental condition. The last two rows show the mean and standard deviation (std) of the accuracies.

Finally, we performed a Wilcoxon rank sum test and no significant differences were observed between the classification accuracies of the HS group and the ALS group ($p = 0.60$). We can say from this result that ALS participants can operate the BMI just as HS would.

3.3. Online Performance and Detection Times

Tables 3, **4** summarize the results obtained in the online evaluation of the proposed BMI. The distribution of the online accuracies, detection times, and ITRs are represented in **Figure 10**. From these results, we can observe that around 46% of the participants achieved an accuracy of 100% in the online tasks. The mean online accuracy was 89.83%, and only three participants obtained accuracies below 75% (HS16, HS17, and ALS5). We can say from this performance evaluation that the implemented BMI decodes the user's intentions effectively in most cases, and users could manipulate the hand-orthosis without much hassle in more complex tasks. However, it is essential to improve the system performance for those users who can not achieve high detection rates.

One way to increase online accuracy is to modify the detection criteria of the interface controller. The number of processed epochs and thresholds for target and non-target classes determine the balance between detection times and classification errors. For instance, if we decrease the non-target class threshold, we can reduce the number of online errors, but it is possible to see higher detection times. Our BMI can customize these parameters for each subject, but for this study, we used the same parameters for all participants.

The average detection time observed in our experiments was 8.54 s, whereas the ITR was 18.13 bit/min. The best and worst times were 2.98 and 13.15 s, and the minimum and maximum ITRs were 2.19 and 52 bit/min. Other studies have reported

TABLE 4 | Average detection times and ITRs obtained in the online evaluation of the proposed BMI.

ID	Average selection time (s)			ITR (bit/min)		
	Without orthosis	With orthosis	Total	Without orthosis	With orthosis	Total
HS1	10.22	11.26	10.87	7.01	8.25	7.72
HS2	10.67	12.47	11.52	14.53	8.78	11.18
HS3	5.43	7.62	6.74	28.56	20.37	23.01
HS4	6.35	6.72	6.54	24.43	23.08	23.73
HS5	8.68	13.56	11.12	17.86	11.44	13.95
HS6	2.94	3.03	2.98	52.83	51.20	52.00
HS7	4.18	4.61	4.40	37.08	33.63	35.27
HS8	4.33	6.49	5.41	35.82	23.91	28.67
HS9	10.56	12.21	11.18	14.69	9.72	12.32
HS10	3.10	4.93	4.02	50.01	31.43	38.60
HS11	10.13	11.05	10.62	10.80	14.04	12.13
HS12	7.61	8.50	8.04	16.92	12.87	14.75
HS13	13.41	12.34	13.14	6.93	7.52	7.07
HS14	3.57	5.94	4.76	43.45	26.10	32.61
HS15	9.54	11.19	10.40	6.84	7.00	6.89
HS16	11.28	15.50	12.84	4.75	2.47	3.67
HS17	7.51	12.61	9.91	3.39	1.45	2.19
HS18	13.84	10.40	13.15	7.91	5.15	7.06
ALS1	NA	9.80	9.80	NA	9.48	9.48
ALS2	NA	8.72	8.72	NA	8.21	8.21
ALS3	NA	7.40	7.40	NA	20.96	20.96
ALS4	NA	10.12	10.12	NA	15.33	15.33
ALS5	9.56	14.83	11.47	8.19	3.61	5.91
ALS6	4.39	5.72	5.06	35.31	27.10	30.66
ALS7	4.84	7.19	5.94	32.05	15.23	21.67
ALS8	6.16	5.56	5.86	25.17	27.90	26.46
Mean	7.65	9.22	8.54	22.02	16.39	18.13
Std	3.26	3.30	3.03	15.34	11.69	12.50

NA indicates absence of validation runs under that experimental condition. The last two rows show the mean and standard deviation (std) of the detection times and ITRs.

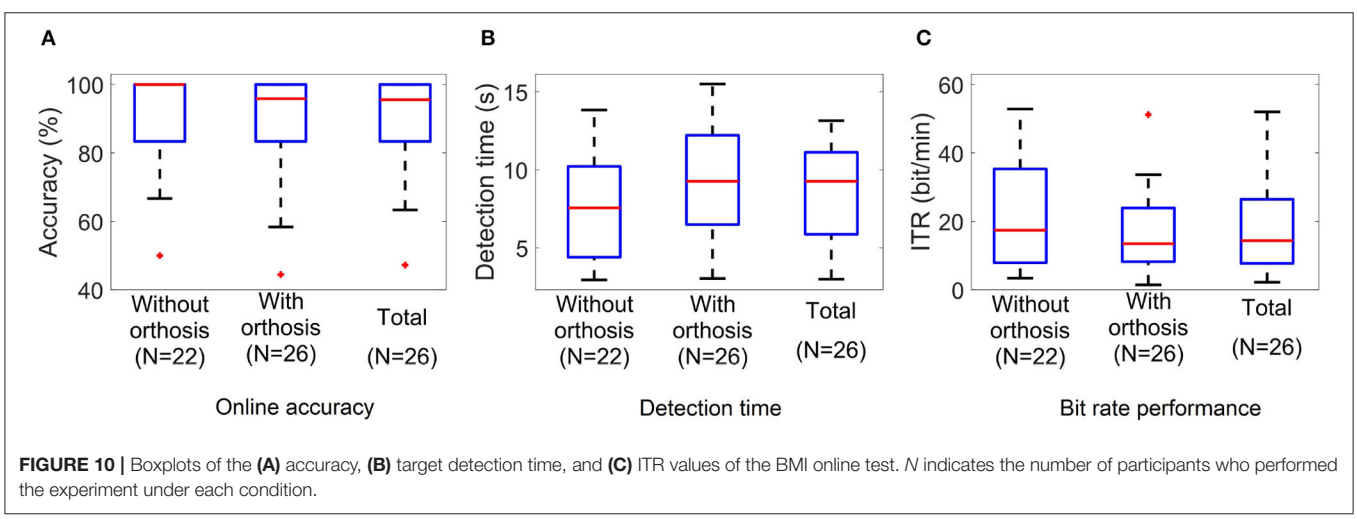


FIGURE 10 | Boxplots of the **(A)** accuracy, **(B)** target detection time, and **(C)** ITR values of the BMI online test. *N* indicates the number of participants who performed the experiment under each condition.

similar results for P300-based BMIs. If we consider that the target population of this technology is people with ALS, these response times are acceptable for many applications such as spellers and smart houses. In the case of a hand-orthosis, it is clear that it is not possible to implement an active fine control for the robotic device. However, users can select complete movements or sequences of actions using our interface. For this reason, we consider that the detection times and ITRs of our system are suitable for the movements or actions contemplated in our BMI.

Performing a Wilcoxon rank sum test to compare the HS group and the ALS group, we do not observe significant differences in any of the three performance metrics studied in this work. Online accuracies ($p = 0.95$), detection times ($p = 0.52$), and ITRs ($p = 0.93$) are similar among groups; consequently, we cannot say the BMI performance is significantly affected by the disease, at least for the disability level of the participants included in this study.

Finally, considering the 22 subjects who performed the experiment without and with orthosis, we carried out a paired t-test to analyze the differences in the system performance between not wearing and wearing the hand-orthosis. While no significant differences in accuracy were found between these two conditions ($t = 1.69$, $df = 21$, $p = 0.1$), the study suggests a significant impact on the detection times ($t = -3.67$, $df = 21$, $p = 0.001$) and ITRs ($t = 3.82$, $df = 21$, $p = 0.001$) produced by the use of the orthosis. These differences may be explained by induced noise mixed with the EEG when the participant wears the hand-orthosis. The linear motors and the power supply of the robotic device produce noise components that can be observed in the electroencephalogram. In this way, the system detects and rejects contaminated epochs more often when the device is turned on and in contact with the user's skin, increasing the detection time. Fortunately, the penalization in the system performance is only 1.57 s, which is not a problem in a P300-based BMI if we consider the typical reaction times of these systems.

4. DISCUSSION

In this work, we presented the development and evaluation of a P300-based BMI coupled with a robotic hand-orthosis. With this system, ALS patients can manipulate each finger of a hand mentally or perform a sequence of movements of one or more fingers. Because the BMI uses the P300 paradigm, the number of possible movements is not limited, and the BMI can provide a range of options for different needs. Our system is able to perform the thumb opposition movement or movements with any combination of fingers, we can also configure the orthosis to be initially closed and perform the extension-flexion of the fingers, the initial position and angular range of the movements of the orthosis can also be controlled, this allows to adapt the system to the individual characteristics of the users (e.g., spasticity, rigidity, level of hand motor impairment), however, for this initial evaluation, we wanted to test the general performance of the interface at the most individual level (single finger movements) and with the most complex movement (all fingers simultaneously), having a total of six possible movements.

In the experiments conducted with HS and ALS patients, we observed event-related activity for the target class in the EEG recordings of all the participants. Additionally, the classification accuracies estimated with cross-validation were above the chance level for all subjects. Finally, in the online tests, both HS and ALS participants were able to control the hand-orthosis with the interface. Only three subjects obtained online accuracies below 75%, and 46% of the study subjects completed all the test runs without errors. These results indicate that our interface can discriminate successfully between target and non-target flashing events, and we can expect that most healthy people and ALS patients with mild to moderate general disability levels (according to the ALSFRS-R scale) are potential users of this assistive technology. After the tests, the users were informally asked about their experience; being the first experience of the subjects with a BMI technology, most of them showed amazement, many of them showed deeply interested and asked about the details of operation and current state of this technology. Some users reported mild eyestrain during the BMI training stage, but all reported feeling physically comfortable during the test.

In this kind of application, it is essential to achieve high accuracies to avoid the user's frustration and increase the chance of acceptance of this technology for daily life use. Although most of the participants obtained low error rates in the conducted experiments, we must find strategies to improve the system performance for users with low classification accuracies. As long as the training data contains observable event-related activity for the target class and the classification model accuracy is above the chance level, we can modify the detection criteria implemented in the interface controller to improve the online performance and adapt the interface to the user's needs. Another possibility would be the modification of the stimulus presentation and the graphical user interface. Some studies have suggested that variations in the visual stimuli characteristics produce variations in the ERPs waveforms, and thus an impact on the BMI performance (Speier et al., 2017; Li et al., 2020).

To our knowledge, this is the first report of a non-invasive P300-based system with multiple possible selections coupled with a robotic hand-orthosis that has been tested with ALS patients. Despite there are previous recent reports of P300-BMIs to control hand-orthosis or artificial hands (Stan et al., 2015; Syrov et al., 2019), these systems were tested only with healthy people, and consider applications mainly for stroke survivors. Stan et al. (2015) presented a system where a hand-orthosis is controlled through a P300-based BMI; however, the system contains only three possible selections (turn on, close, and open orthosis) that allow the flexion-extension of the five fingers simultaneously, while our system allows the passive flexion-extension of a single finger at a time. The evaluation of these fine motor movements is particularly important in ALS patients since this disease is directly associated with the degeneration of the corticospinal tract (Sarica et al., 2017), which is involved in fine digital movements (Levine et al., 2012). Syrov et al. (2019) developed a P300-BMI approach to control each finger of

an wired, artificial phantom hand which does not perform the passive flexion-extension of the users fingers. In their system, the visual stimuli are shown through LEDs placed directly on the fingers of the artificial hand; this configuration, in addition to the absence of wireless communication to the robotic hand, could bring additional difficulties to test the system with ALS patients due to their motor limitations. On the other hand, Gull et al. (2018) proposed a prototype intended to be used with ALS patients that includes a BMI and a robotic glove to assist hand grasping; nevertheless, the robotic glove (Nilsson et al., 2012) covers only three fingers (thumb, middle, and ring), and the implementation of the BMI paradigm, glove control, and clinical tests were reported inconclusive.

The datasets of each participant collected in this study are publicly available with the idea of contributing to the development of new processing and classification methods for BMI systems. The inclusion of datasets of ALS participants increases the available information containing EEG recordings for BMI purposes and facilitates the improvement of BMI-based tools for patients. Furthermore, the ERPs could be used to investigate potential electrophysiological biomarkers of ALS (McCane et al., 2015; Lange et al., 2016), which would help to understanding the neurodegenerative mechanisms of the disease.

In conclusion, the results presented in this work show the capability of our mind-controlled hand-orthosis to be used with no need of adaptations for ALS patients with moderate level of disability. Future work will focus on increasing the sample size of ALS users and investigating the effect of longitudinal use of the system on patients. We will also modify the available options of the interface to test more realistic scenarios. Our system could represent the basis for developing more practical tools, such as a portable orthosis that responds to other biosignals in addition to the EEG and that is adaptable to the degree of disability of the users. Our system could also be modified to communicate with other wireless systems (e.g., smart homes).

For this initial evaluation, we tested our system's effectiveness and efficiency in terms of accuracy and ITR, respectively; for our future work, we will adopt an user-centered design (UCD) approach (Liberati et al., 2015; Schettini et al., 2015; Riccio et al., 2016; Kübler et al., 2020) and include the evaluation of satisfaction by consulting and registering the opinion of primary (ALS patients) and secondary (caregivers) end-users through formal interviews. Feedback from patients and their caregivers will help to develop a more customizable system according to the individual characteristics and needs of each user. The UCD approach will also help us to properly identify and correct the present limitations in order to improve the usability of our system in daily life.

DATA AVAILABILITY STATEMENT

The datasets generated and analyzed for this study are available upon request to the corresponding author.

ETHICS STATEMENT

All participants volunteered for the study and provided informed consent before the experimental sessions.

AUTHOR CONTRIBUTIONS

JD, OM-M, JG, RC, HM, and JMA participated in the study design, experiments, and manuscript writing. JD and OM-M designed and implemented the brain-machine interface. JD, OM-M, and JA performed the acquisition and analysis of data. JD recruited the healthy participants. RC and HM selected the amyotrophic lateral sclerosis patients.

FUNDING

This research has been funded by the National Council of Science and Technology of Mexico (CONACyT) through grant PN2015-873.

REFERENCES

Achanccaray, D., Flores, C., Fonseca, C., and Andreu-Perez, J. (2017). "A P300-based brain computer interface for smart home interaction through an ANFIS ensemble," in *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (Naples), 1–5. doi: 10.1109/FUZZ-IEEE.2017.8015770

Aggogeri, F., Mikolajczyk, T., and O'Kane, J. (2019). Robotics for rehabilitation of hand movement in stroke survivors. *Adv. Mech. Eng.* 11:168781401984192. doi: 10.1177/1687814019841921

Arrichiello, F., Di Lillo, P., Di Vito, D., Antonelli, G., and Chiaverini, S. (2017). "Assistive robot operated via p300-based brain computer interface," in *2017 IEEE International Conference on Robotics and Automation (ICRA)* (Singapore), 6032–6037. doi: 10.1109/ICRA.2017.7989714

Aydin, E. A., Bay, O. F., and Guler, I. (2018). P300-based asynchronous brain computer interface for environmental control system. *IEEE J. Biomed. Health Informatics* 22, 653–663. doi: 10.1109/JBHI.2017.2690801

Berrar, D. (2019). "Cross-validation," in *Encyclopedia of Bioinformatics and Computational Biology*, eds S. Ranganathan, M. Gribskov, K. Nakai, and C. Schönbach (Oxford: Academic Press), 542–545. doi: 10.1016/B978-0-12-809633-8.20349-X

Bowman, A., and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis*. Number 18 in Oxford statistical science series. Oxford: Clarendon Press.

Cattan, G. H., Andreev, A., Mendoza, C., and Congedo, M. (2019). A comparison of mobile VR display running on an ordinary smartphone with standard pc display for p300-bci stimulus presentation. *IEEE Trans. Games* 2019:2957963. doi: 10.1109/TG.2019.2957963

Chen, L., Jin, J., Daly, I., Zhang, Y., Wang, X., and Cichocki, A. (2016). Exploring combinations of different color and facial expression stimuli for gaze-independent BCIs. *Front. Comput. Neurosci.* 10:5. doi: 10.3389/fncom.2016.00005

Chen, L., Jin, J., Zhang, Y., Wang, X., and Cichocki, A. (2015). A survey of the dummy face and human face stimuli used in BCI paradigm. *J. Neurosci. Methods* 239, 18–27. doi: 10.1016/j.jneumeth.2014.10.002

Cho, J., Jeong, J., Shim, K., Kim, D., and Lee, S. (2018). "Classification of hand motions within EEG signals for non-invasive BCI-based robot hand control,"

in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (Miyazaki), 515–518. doi: 10.1109/SMC.2018.00097

De Venuto, D., Annese, V. F., and Mezzina, G. (2017). "An embedded system remotely driving mechanical devices by P300 brain activity," in *Design, Automation Test in Europe Conference Exhibition (DATE)* (Lausanne), *2017*, 1014–1019. doi: 10.23919/DATE.2017.7927139

De Venuto, D., Annese, V. F., and Mezzina, G. (2018). Real-time P300-based bci in mechatronic control by using a multi-dimensional approach. *IET Softw.* 12, 418–424. doi: 10.1049/iet-sen.2017.0340

Deligani, R. J., Hosni, S. I., Vaughan, T. M., McCane, L. M., Zeitlin, D. J., McFarland, D. J., et al. (2019). "Neural alterations during use of a P300-based BCI by individuals with amyotrophic lateral sclerosis*," in *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)* (San Francisco, CA), 899–902. doi: 10.1109/NER.2019.8717044

Flores, C., Fonseca, C., Achanccaray, D., and Andreu-Perez, J. (2018). "Performance evaluation of a P300 brain-computer interface using a kernel extreme learning machine classifier," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (Miyazaki), 3715–3719. doi: 10.1109/SMC.2018.00629

Garakani, G., Ghane, H., and Menhaj, M. B. (2019). Control of a 2-DOF robotic arm using a P300-based brain-computer interface. *AUT J. Model. Simul.* doi: 10.22060/miscj.2019.15569.5136

Gondesen, F., Marx, M., and Kycler, A. (2019). "A shoulder-surfing resistant image-based authentication scheme with a brain-computer interface," in *2019 International Conference on Cyberworlds (CW)* (Kyoto), 336–343. doi: 10.1109/CW.2019.00061

Good, P. I. (2006). *Resampling Methods: A Practical Guide to Data Analysis*. Boston, MA: Birkhäuser.

Gull, M. A., Bai, S., Mrachacz-Kersting, N., and Blicher, J. (2018). "Wexo: Smart wheelchair exoskeleton for ALS patients," in *Proceedings of the 12th International Convention on Rehabilitation Engineering and Assistive Technology, i-CREATe 2018* (Midview City: Singapore Therapeutic, Assistive & Rehabilitative Technologies (START) Centre), 97–100.

Guy, V., Soriani, M.-H., Bruno, M., Papadopoulo, T., Desnuelle, C., and Clerc, M. (2018). Brain computer interface with the P300 speller: usability for disabled people with amyotrophic lateral sclerosis. *Ann. Phys. Rehabil. Med.* 61, 5–11. doi: 10.1016/j.rehab.2017.09.004

Ho, N. S. K., Tong, K. Y., Hu, X. L., Fung, K. L., Wei, X. J., Rong, W., et al. (2011). "An EMG-driven exoskeleton hand robotic training device on chronic stroke subjects: task training system for stroke rehabilitation," in *2011 IEEE International Conference on Rehabilitation Robotics* (Zurich), 1–5. doi: 10.1109/ICORR.2011.5975340

Hwang, H.-J., Kim, S., Choi, S., and Im, C.-H. (2013). EEG-based brain-computer interfaces: a thorough literature survey. *Int. J. Human. Comput. Interact.* 29, 814–826. doi: 10.1080/10447318.2013.780869

Ivy, C. C., Smith, S. M., and Materi, M. M. (2014). Upper extremity orthoses use in amyotrophic lateral sclerosis/motor neuron disease: three case reports. *HAND* 9, 543–550. doi: 10.1007/s11552-014-9626-x

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2015). *An Introduction to Statistical Learning: With Applications in R*. New York, NY: Springer Publishing Company, Incorporated.

Kleih, S. C., Gottschalt, L., Teichlein, E., and Weilbach, F. X. (2016). Toward a P300 based brain-computer interface for aphasia rehabilitation after stroke: presentation of theoretical considerations and a pilot feasibility study. *Front. Hum. Neurosci.* 10:547. doi: 10.3389/fnhum.2016.00547

Kübler, A., Nijboer, F., and Kleih, S. (2020). "Chapter 26 - hearing the needs of clinical users," in *Brain-Computer Interfaces, Vol. 168 of Handbook of Clinical Neurology*, eds N. F. Ramsey and J. del R. Millan (Elsevier), 353–368. doi: 10.1016/B978-0-444-63934-9.00026-3

Lange, F., Lange, C., Joop, M., Seer, C., Dengler, R., Kopp, B., et al. (2016). Neural correlates of cognitive set shifting in amyotrophic lateral sclerosis. *Clin. Neurophysiol.* 127, 3537–3545. doi: 10.1016/j.clinph.2016.09.019

Levine, A., Lewallen, K., and Pfaff, S. (2012). Spatial organization of cortical and spinal neurons controlling motor behavior. *Curr. Opin. Neurobiol.* 22, 812–821. doi: 10.1016/j.conb.2012.07.002

Li, S., Jin, J., Daly, I., Zuo, C., Wang, X., and Cichocki, A. (2020). Comparison of the ERP-based BCI performance among chromatic (RGB) semitransparent face patterns. *Front. Neurosci.* 14:54. doi: 10.3389/fnins.2020.00054

Liberati, G., Pizzimenti, A., Simione, L., Riccio, A., Schettini, F., Inghilleri, M., et al. (2015). Developing brain-computer interfaces from a user-centered perspective: assessing the needs of persons with amyotrophic lateral sclerosis, caregivers, and professionals. *Appl. Ergon.* 50, 139–146. doi: 10.1016/j.apergo.2015.03.012

Lotte, F., and Guan, C. (2009). "An efficient p300-based brain-computer interface with minimal calibration time," in *Assistive Machine Learning for People With Disabilities Symposium (NIPS'09 Symposium)* (Whistler).

Martínez, H. R., Figueroa-Sánchez, J. A., Cantú-Martínez, L., Caraza, R., de la Maza, M., Escamilla-Garza, J. M., et al. (2020). A multidisciplinary clinic for amyotrophic lateral sclerosis patients in northeast mexico. *Rev. Mex. Neuroci.* 21, 66–70. doi: 10.24875/RMN.19000144

Masud, U., Baig, M. I., Akram, F., and Kim, T. (2017). "A P300 brain computer interface based intelligent home control system using a random forest classifier," in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)* (Honolulu), 1–5. doi: 10.1109/SSCI.2017.8285449

McCane, L., Heckman, S., McFarland, D., Townsend, G., Mak, J., Sellers, E., et al. (2015). P300-based brain-computer interface (BCI) event-related potentials (ERPs): people with amyotrophic lateral sclerosis (ALS) vs. age-matched controls. *Clin. Neurophysiol.* 126, 2124–2131. doi: 10.1016/j.clinph.2015.01.013

McFarland, D. J. (2020). Brain-computer interfaces for amyotrophic lateral sclerosis. *Muscle Nerve* 61, 702–707. doi: 10.1002/mus.26828

Mendoza-Montoya, O. (2017). *Development of a hybrid brain-computer interface for autonomous systems* (Ph.D. thesis). Freie Universitat Berlin, Berlin, Germany.

Murphy, M. D., Guggenmos, D. J., Bundy, D. T., and Nudo, R. J. (2016). Current challenges facing the translation of brain computer interfaces from preclinical trials to use in human patients. *Front. Cell. Neurosci.* 9:497. doi: 10.3389/fncel.2015.00497

Nilsson, M., Ingvast, J., Wikander, J., and von Holst, H. (2012). "The soft extra muscle system for improving the grasping capability in neurological rehabilitation," in *2012 IEEE-EMBS Conference on Biomedical Engineering and Sciences* (Langkawi), 412–417. doi: 10.1109/IECBES.2012.6498090

Nurseitov, D., Serekov, A., Shintemirov, A., and Abibullaev, B. (2017). "Design and evaluation of a P300-ERP based BCI system for real-time control of a mobile robot," in *2017 5th International Winter Conference on Brain-Computer Interface (BCI)* (Jeongseon), 115–120. doi: 10.1109/IWW-BCI.2017.7858177

Okahara, Y., Takano, K., Komori, T., Nagao, M., Iwadate, Y., and Kansaku, K. (2017). Operation of a P300-based brain-computer interface by patients with spinocerebellar ataxia. *Clin. Neurophysiol. Pract.* 2, 147–153. doi: 10.1016/j.cnp.2017.06.004

Poletti, B., Carelli, L., Solca, F., Lafronza, A., Pedroli, E., Faini, A., et al. (2016). Cognitive assessment in amyotrophic lateral sclerosis by means of P300-brain computer interface: a preliminary study. *Amyotr. Lateral Scler. Frontotemp. Degener.* 17, 473–481. doi: 10.1080/21678421.2016.1181182

Riccio, A., Pichiorri, F., Schettini, F., Toppi, J., Risetti, M., Formisano, R., et al. (2016). "Chapter 12 - interfacing brain with computer to improve communication and rehabilitation after brain damage," in *Brain-Computer Interfaces: Lab Experiments to Real-World Applications, Vol. 228 of Progress in Brain Research*, ed D. Coyle (Amsterdam: Elsevier), 357–387. doi: 10.1016/bs.pbr.2016.04.018

Sarica, A., Cerasa, A., Valentino, P., Yeatman, J., Trotta, M., Barone, S., et al. (2017). The corticospinal tract profile in amyotrophic lateral sclerosis. *Hum. Brain Mapp.* 38, 727–739. doi: 10.1002/hbm.23412

Schettini, F., Riccio, A., Simione, L., Liberati, G., Caruso, M., Frasca, V., et al. (2015). Assistive device with conventional, alternative, and brain-computer interface inputs to enhance interaction with the environment for people with amyotrophic lateral sclerosis: a feasibility and usability study. *Arch. Phys. Med. Rehabil.* 96(3 Suppl.), S46–S53. doi: 10.1016/j.apmr.2014.05.027

Shahriari, Y., Vaughan, T. M., McCane, L. M., Allison, B. Z., Wolpaw, J. R., and Krusienski, D. J. (2019). An exploration of BCI performance variations in people with amyotrophic lateral sclerosis using longitudinal EEG data. *J. Neural Eng.* 16:056031. doi: 10.1088/1741-2552/ab22ea

Speier, W., Deshpande, A., Cui, L., Chandravadia, N., Roberts, D., and Pouratian, N. (2017). A comparison of stimulus types in online classification of the P300 speller using language models. *PLoS ONE* 12:e175382. doi: 10.1371/journal.pone.0175382

Spüler, M., Walter, A., Rosenstiel, W., and Bogdan, M. (2014). Spatial filtering based on canonical correlation analysis for classification of evoked or event-related potentials in EEG data. *IEEE Trans. Neural Syst. Rehabil. Eng.* 22, 1097–1103. doi: 10.1109/TNSRE.2013.2290870

Stan, A., Irimia, D. C., Botezatu, N. A., and Lupu, R. G. (2015). "Controlling a hand orthosis by means of P300-based brain computer interface," in *2015 E-Health and Bioengineering Conference (EHB)* (Iasi), 1–4. doi: 10.1109/EHB.2015.7391389

Syrov, N., Novichikhina, K., Kirayanov, D., Gordleeva, S., and Kaplan, A. (2019). The changes of corticospinal excitability during the control of artificial hand through the brain-computer interface based on the P300 component of visual evoked potential. *Hum. Physiol.* 45, 152–157. doi: 10.1134/S0362119719020117

Tanaka, K., Horaiya, K., Akagi, J., and Kihoin, N. (2013). Timely manner application of hand orthoses to patients with amyotrophic lateral sclerosis: a case report. *Prosthet. Orthot. Int.* 38, 239–242. doi: 10.1177/0309364613489334

Tang, J., Zhou, Z., and Liu, Y. (2017). "A 3D visual stimuli based p300 brain-computer interface: for a robotic arm control," in *Proceedings of the 2017 International Conference on Artificial Intelligence, Automation and Control Technologies, AIACT'17* (New York, NY: Association for Computing Machinery), 1–6. doi: 10.1145/3080845.3080863

Tyagi, A., and Nehra, V. (2017). A comparison of feature extraction and dimensionality reduction techniques for EEG-based BCI system. *IUP J. Comput. Sci.* 11, 51–66.

Vaughan, T. M. (2020). "Chapter 4 - brain-computer interfaces for people with amyotrophic lateral sclerosis," in *Brain-Computer Interfaces, Vol. 168 of Handbook of Clinical Neurology*, eds N. F. Ramsey and R. del Millán (Amsterdam: Elsevier), 33–38. doi: 10.1016/B978-0-444-63934-9.00004-4

Venuto, D. D., Annese, V. F., Mezzina, G., Ruta, M., and Sciascio, E. D. (2016). "Brain-computer interface using P300: a gaming approach for neurocognitive impairment diagnosis," in *2016 IEEE International High Level Design Validation and Test Workshop (HLDVT)* (Santa Cruz, CA), 93–99.

Wang, W., and Chakraborty, G. (2017). "Probes minimization still maintaining high accuracy to classify target stimuli P300," in *2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST)* (Taichung), 13–17. doi: 10.1109/ICAwST.2017.8256431

WMA (2013). World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. *JAMA* 310, 2191–2194. doi: 10.1001/jama.2013.281053

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 767–791. doi: 10.1016/S1388-2457(02)00057-3

Won, K., Kwon, M., Lee, S., Jang, S., Lee, J., Ahn, M., et al. (2018). "Seeking rsvp task features correlated with P300 speller performance," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (Miyazaki), 1138–1143. doi: 10.1109/SMC.2018.00201

Xiao, X., Xu, M., Jin, J., Wang, Y., Jung, T., and Ming, D. (2019). Discriminative canonical pattern matching for single-trial classification of erp components. *IEEE Trans. Biomed. Eng.* 67, 2266–2275. doi: 10.1109/TBME.2019.2958641

Yu, M., Kaongoen, N., and Jo, S. (2016). "P300-BCI-based authentication system," in *2016 4th International Winter Conference on Brain-Computer Interface (BCI)* (Gangwon-do), 1–4. doi: 10.1109/IWW-BCI.2016.7457443

Zhumadilova, A., Tokmurzina, D., Kuderbekov, A., and Abibullaev, B. (2017). "Design and evaluation of a P300 visual brain-computer interface speller in cyrillic characters," in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (Lisbon), 1006–1011. doi: 10.1109/ROMAN.2017.8172426

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 01 December 2020
doi: 10.3389/fnins.2020.591777

Impact of Stimulus Features on the Performance of a Gaze-Independent Brain-Computer Interface Based on Covert Spatial Attention Shifts

Christoph Reichert[1,2,3], Igor Fabian Tellez Ceja[4], Catherine M. Sweeney-Reed[2,5], Hans-Jochen Heinze[1,2,5], Hermann Hinrichs[1,2,3,5] and Stefan Dürschmid[1,5]*

[1] Department of Behavioral Neurology, Leibniz Institute for Neurobiology, Magdeburg, Germany, [2] Center for Behavioral Brain Sciences, Magdeburg, Germany, [3] Research Campus STIMULATE, Magdeburg, Germany, [4] Institute for Medical Engineering, Otto-von-Guericke University, Magdeburg, Germany, [5] Department of Neurology, Otto-von-Guericke University, Magdeburg, Germany

OPEN ACCESS

Edited by:
Mikhail Lebedev, Duke University, United States

Reviewed by:
Jing Jin, East China University of Science and Technology, China
Yijun Wang, Institute of Semiconductors (CAS), China

***Correspondence:**
Christoph Reichert
christoph.reichert@lin-magdeburg.de

Specialty section:
This article was submitted to Neural Technology, a section of the journal Frontiers in Neuroscience

Received: *05 August 2020*
Accepted: *10 November 2020*
Published: *01 December 2020*

Citation:
Reichert C, Tellez Ceja IF, Sweeney-Reed CM, Heinze H-J, Hinrichs H and Dürschmid S (2020) Impact of Stimulus Features on the Performance of a Gaze-Independent Brain-Computer Interface Based on Covert Spatial Attention Shifts. Front. Neurosci. 14:591777. doi: 10.3389/fnins.2020.591777

Regaining communication abilities in patients who are unable to speak or move is one of the main goals in decoding brain waves for brain-computer interface (BCI) control. Many BCI approaches designed for communication rely on attention to visual stimuli, commonly applying an oddball paradigm, and require both eye movements and adequate visual acuity. These abilities may, however, be absent in patients who depend on BCI communication. We have therefore developed a response-based communication BCI, which is independent of gaze shifts but utilizes covert shifts of attention to the left or right visual field. We recorded the electroencephalogram (EEG) from 29 channels and coregistered the vertical and horizontal electrooculogram. Data-driven decoding of small attention-based differences between the hemispheres, also known as N2pc, was performed using 14 posterior channels, which are expected to reflect correlates of visual spatial attention. Eighteen healthy participants responded to 120 statements by covertly directing attention to one of two colored symbols (green and red crosses for "yes" and "no," respectively), presented in the user's left and right visual field, respectively, while maintaining central gaze fixation. On average across participants, 88.5% (std: 7.8%) of responses were correctly decoded online. In order to investigate the potential influence of stimulus features on accuracy, we presented the symbols with different visual angles, by altering symbol size and eccentricity. The offline analysis revealed that stimulus features have a minimal impact on the controllability of the BCI. Hence, we show with our novel approach that spatial attention to a colored symbol is a robust method with which to control a BCI, which has the potential to support severely paralyzed people with impaired eye movements and low visual acuity in communicating with their environment.

Keywords: visual spatial attention, brain-computer interface, stimulus features, N2pc, canonical correlation analysis, gaze-independent, BCI

INTRODUCTION

A brain-computer interface (BCI) that can be controlled independently of gaze shifts could constitute a helpful assistive device for persons who suffer from severe neurological disorders. However, most developments in the field of BCI presume that the users can move their eyes. One of the most extensively studied brain signals is the steady-state-visual-evoked potential (SSVEP; Müller-Putz et al., 2005; Lin et al., 2007; Vialatte et al., 2010; Zhu et al., 2010), because its signal-to-noise ratio is relatively high, and low training effort is required to set up the decoder. It is commonly used in overt BCI control, since during covert attention paradigms, the behavioral performance, the SSVEP amplitude and BCI accuracy are comparatively reduced (Kelly et al., 2004; Walter et al., 2012). Another prominent example is the matrix speller, which was initially introduced by Farwell and Donchin (1988) and utilizes the P300 response to detect the time point at which the stimulus is presented at the target symbol location, on which the users' attention is focused. Overt attention has also been shown to enable more reliable control than covert attention using matrix spellers (Brunner et al., 2010; Treder and Blankertz, 2010). Reliability is greater during overt compared with covert attention due to the additional modulation of early visual event-related potential (ERP) components according to the focus of attention (Treder and Blankertz, 2010; Frenzel et al., 2011), which is deemed to result from greater central than peripheral visual acuity. Hence, paradigms have been developed, which make use of more centrally located presentations (Treder et al., 2011b). For example, rapid serial visual presentation has been applied to detect a target in a series of rapidly presented symbols (Acqualagna et al., 2010; Lin et al., 2018). The disadvantage of this paradigm is that target presentations could be missed due to the attentional blink (Raymond et al., 1992). Moreover, in general, vision-based BCIs require good visual acuity, even if they are gaze-independent, but potential users frequently suffer from impaired vision (Halder et al., 2016). For this reason, auditory (Kübler et al., 2009; Halder et al., 2010, 2016; Hill et al., 2014) and tactile variants (Brouwer and van Erp, 2010; Jin et al., 2020) of the oddball paradigm have been investigated, with the finding that they provide inferior performance compared to BCIs based on visual stimuli (Severens et al., 2014).

In summary, a great deal of research into BCIs is dependent upon participants' ability to execute eye movements. This requirement, however, largely neglects the fact that the main aim in BCI development is to provide a means of communication and control for patients in whom the ability to execute eye movements is impaired. Recently, it has been shown that spatial attention to peripherally presented colored stimuli permits reliable, gaze-independent control of a four class BCI (Reichert et al., 2020a). The paradigm takes advantage of the fact that shifts in attention to targets that pop up in the periphery of the visual field evoke slight interhemispheric differences in brain activity, depending on the side where the target was presented. This phenomenon has been intensively investigated in visual search experiments, where targets were presented in a search display among distractors, e.g., (Heinze et al., 1990; Luck and Hillyard, 1994a; Luck and Ford, 1998). Specifically, it has been found that in parieto-occipital regions contralateral to the presented target, 180–300 ms after onset, a stronger negative deflection compared to ipsilateral sites can be measured with the electroencephalogram (EEG). This component is known as the N2pc (Luck and Hillyard, 1994b) and is assumed to reflect the attentional selection of target features (Eimer, 1996). The fact that paying attention to simple features like color evokes spatially different ERPs depending on the visual hemifield where it was presented, suggests that the N2pc may be suited to gaze-independent BCI control. However, the potential advantages of using shifts in spatial attention have not yet been systematically evaluated for use in BCI control. The ability to classify several positions of peripherally presented targets has been evaluated using alpha activity (van Gerven and Jensen, 2009; Treder et al., 2011a) and ERPs (Fahrenfort et al., 2017). Classification of hemispheric differences, depending on the hemifield in which the target was presented, has been successfully applied for target detection in aerial images (Matran-Fernandez and Poli, 2017), for the detection of the tilt of Gabor patches (Xu et al., 2016) as well as in visual search for colored digits (Awni et al., 2013) and circles (Tian et al., 2019). While data in these studies were analyzed offline, to our knowledge, only one study has implemented a gaze-independent closed-loop BCI based on N2pc detection (Reichert et al., 2020a), where participants performed a two-dimensional navigation task. Here we extend this initial work to evaluate how stimulus size and eccentricity modulate the N2pc, which could alter the accuracy of the BCI. Specifically, we implemented a BCI for binary communication, suitable for responding to dichotomous questions. We hypothesize that hemispheric differences related to spatial attention are largely independent of stimulus size. This would permit the BCI to be operated with relatively large stimuli such that patients with low visual acuity can control the system. While the potential role of distractors in the composition of the N2pc has been investigated in a number of studies, with conclusions remaining controversial (Luck et al., 1997; Hopf et al., 2002; Hickey et al., 2009; Mazza et al., 2009), there have been no systematic investigations of the impact of symbol size and only one recent study exploring the impact of target eccentricity on the N2pc component (Papaioannou and Luck, 2020). However, BCI accuracy might depend on stimulus features, as such a dependency has been revealed using BCIs based on the P300 potential. For instance, accuracy was increased when faces were presented as stimuli as opposed to character flashes or meaningless images (Kaufmann et al., 2011), when luminance and chromatic features were combined (Takano et al., 2009) and when 3D stereo visual stimuli were presented as opposed to 2D stimuli (Qu et al., 2018). In contrast, symbol size and inter-symbol distance appears to have no general effect on the performance (Salvaris and Sepulveda, 2009). In the BCI experiment presented here, we varied symbol sizes and eccentricities to investigate whether such stimulus features have an impact on classification accuracy, and if so, to determine the optimal set of stimulus features to prevent poor performance due to inappropriate parameter choices in future studies.

MATERIALS AND METHODS

Participants and Recordings

Eighteen healthy participants (10 female, 19 to 38 years, mean age: 27 years) took part in the study. All participants had normal or corrected to normal vision and reported no neurological impairment. They gave written informed consent and were paid for their participation. The study was approved by the Ethics Committee of the Otto-von-Guericke University, Magdeburg.

Participants were seated in an acoustically shielded and dimly lit cabin and viewed a 24" display (ASUS VG248QE) from a distance of 70 cm. Visual stimuli were registered by a photodiode to synchronize screen events with the EEG. The EEG was recorded from 29 Ag/AgCl electrodes, placed at standard positions of an extended 10–20 montage, using a BrainAmp DC Amplifier (Brain Products GmbH, Germany). Electrode measurements were referenced against the right mastoid and sampled at 250 Hz. Furthermore, the vertical and horizontal EOG (hEOG) was recorded simultaneously to register eye movements. Parallel to the recordings, EEG signals were transferred through TCP/IP to the BCI client.

EOG Calibration

Before the experiment started, we recorded the EOG while participants were presented with a cross which they were asked to track with their gaze, and which changed its position every 1,250 ms. The position displacement relative to the center varied from 1 degree to 7 degrees horizontally and in 30% of trials we additionally displaced the cross by 2 degrees of visual angle vertically. Three times the cross was replaced by a circle, and participants were asked to perform an eye blink immediately. In total, 40 gaze shifts and three blinks were performed in an unpredictable order, resulting in approximately 1 min of EOG calibration. This procedure provided us with calibration data which characterize the strength of EOG signals as a function of gaze shift angle. We used these data to evaluate the degree of unintentional eye movement during BCI control.

Stimulus and Task

Participants were asked to respond to yes/no questions or statements by shifting their attention to a green +-cross to respond with "yes" or to a red × -cross to respond with "no" (see **Figure 1**). The first 96 questions and statements could be objectively answered with "yes" or "no," e.g., "Is Berlin a city?". To reduce the probability that there is a bias toward one particular answer, each question or statement also had a counterpart (e.g., "Is Berlin a continent?"), such that the numbers of expected "yes" and "no" answers were balanced. The last 24 questions could only be answered by the participant subjectively (e.g., "Are you a vegetarian?"), which constituted a demonstration of real-world application. Note that a correct answer was not relevant for the BCI. The BCI only evaluated the attentional shift, as decoded from the EEG and fed the result back to the participants. In turn, participants evaluated whether their intended response and the BCI feedback matched. A button press with the index finger indicated correct BCI feedback and a button press with the middle finger indicated that the feedback was not correct. Sixteen participants were native German speakers and were presented with questions in German. Two participants were not German but fluent in English, and we presented the same questions in English.

Each trial started with presentation of a question. The participants had time to select the response and direct their gaze on the upper or lower but central fixation cross corresponding to the answer until they pressed a button to start the stimulus sequence. The differently colored fixation crosses were presented to help the participant to keep the target in mind during the entire trial. For example, if the participants decided to answer with yes, they directed their gaze to the green central fixation cross and focused their attention on the green +-cross, which was presented randomly left or right, during the whole stimulus sequence that followed. A single stimulus sequence comprised a series of ten stimuli, which was found previously to provide a good trade-off between stimulation time and accuracy (Reichert et al., 2020a). A single stimulus display thus consisted of a red × -cross presented in the left or right visual field and a green +-cross presented at the opposite visual field. The position at which the cross symbols appeared was surrounded by 8 gray dots, which were uninterruptedly presented throughout the whole trial to indicate the position where the stimulus would appear. In each stimulus display, we presented both a red and a green cross, one to the left and one to the right. Across the stimulus sequence, the colors were pseudo-randomly allocated to visual fields, with the restriction that the number of left/right presentations was balanced for both colors and that the same color was presented in the same visual field in a maximum of three consecutive stimulus displays. Each stimulus display was presented for 250 ms with a stimulus onset asynchrony of 850 ms, jittered by 0–250 ms. For an example stimulus sequence see the **Supplementary Material**. The horizontal position α and size of the stimuli φ was constant within a trial but varied between trials. We used four different φ levels, i.e., four symbol sizes which we define in visual angles ($\varphi_1 = 0.45°$, $\varphi_2 = 0.90°$, $\varphi_3 = 1.36°$, and $\varphi_4 = 1.81°$) and five different α levels, i.e., five eccentricities which we define in visual angles ($\alpha_1 = 4°$, $\alpha_2 = 5.5°$, $\alpha_3 = 7°$, $\alpha_4 = 8.5°$, and $\alpha_5 = 10°$). See also **Figure 2** for a definition of parameters α and φ. We combined parameters φ_j with α_j and with α_{j+1}, $j = 1...4$, such that 8 parameter pairs were tested. Each parameter set was applied in random order, three times per block, resulting in 24 trials per block. After the stimulus sequence was presented, feedback "yes" or "no" as decoded with the BCI, was presented. The participants confirmed the correctness of the feedback as described above. The resulting response was considered the ground truth, which we used to train and evaluate the BCI.

The first two blocks were conducted to acquire data to train the classifier. Therefore, we did not present questions in those blocks but instructed the participant to shift attention to the green symbol in the first block and to the red symbol in the second block. Afterward, we presented five blocks with 24 questions or statements each, except for one participant who performed only three question blocks due to technical issues. The classifier was initially trained with data from the first two blocks and retrained after each trial.

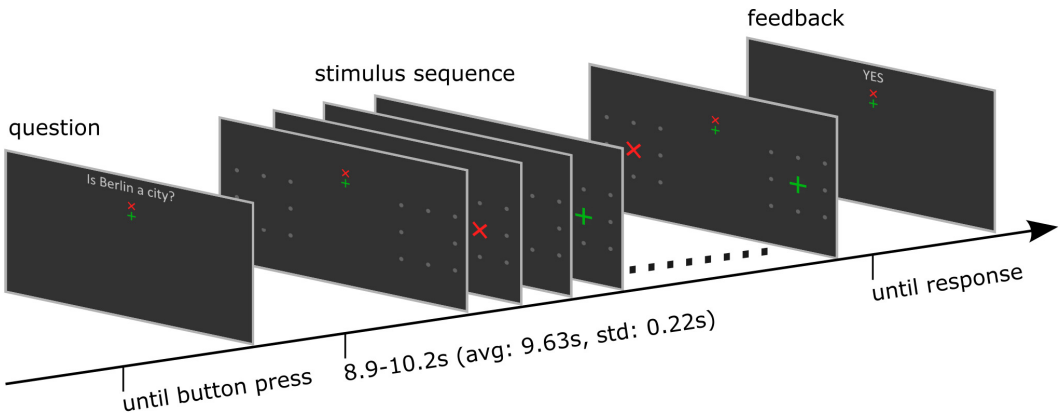


FIGURE 1 | Structure of one trial. The trial started with a dichotomous question. The participant started the stimulation sequence by button press. After the stimulus presentation was finished, the BCI determined the target symbol, and the corresponding feedback was presented. Finally, participants indicated its correctness by responding with a further button press. See also the **Supplementary Material** for a video demonstration.

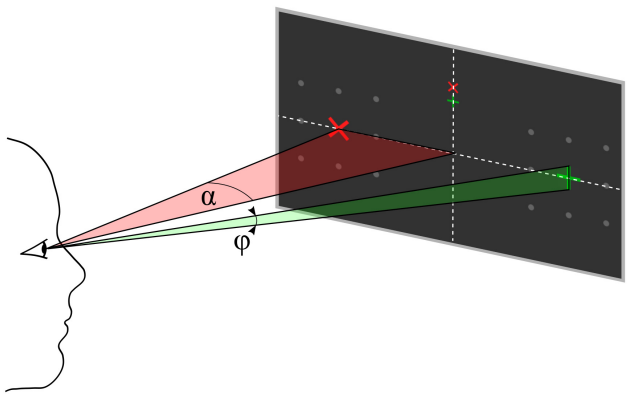


FIGURE 2 | Parameter definition for eccentricity α and symbol size φ, both defined as visual angles. The participant's gaze is directed to one of the upper central crosses.

Processing of EEG Data

In order to prevent hemispheric differences induced by a unilateral reference electrode, we re-referenced the EEG data to the average of left and right mastoid. Although we recorded 29 channels to provide full head coverage as open data (Reichert et al., 2020b), we used only 14 parieto-occipital channels (O9, O10, CP1, CP2, Pz, P3, P4, P7, P8, PO3, PO4, PO7, PO8, and Oz) to decode the shifts in visual attention. The EEG data corresponding to a stimulus sequence of a trial were cut out according to the start and stop events that we sent as trigger signals to the EEG device before and after presentation of the stimulus sequence. A 4th order zero-phase IIR Butterworth bandpass filter between 1.0 and 12.5 Hz was applied to the data, which were then resampled to 50 Hz. The stimulus onsets were determined from the signal sent by the photodiode. Afterward, the time series data were epoched starting from stimulus onset to 750 ms after stimulus onset. Since this resulted in 38 sampling points involving 14 channels, we can write an epoch as a matrix $X_i \in \mathbb{R}^{38\times14}$. Since the epochs of one trial, represented by a sequence of ten stimuli, refer to the same target item, data of one trial are composed of ten epochs. Epochs in which the green symbol was presented on the left and the red symbol on the right were labeled with $y_i = 1$, while epochs in which the red and green symbols were presented the other way around were labeled with $y_i = -1$.

Decoding Approach

In this section we first describe the estimation of the decoding model, which is performed each time the classifier is trained – online and during the folds of cross validation. Afterward, we describe how to apply this model to unseen data – both to present online feedback and also to decode left-out trials in a cross validation. An implementation of the decoding approach can be found in the publicly available data set (Reichert et al., 2020b).

Model Estimation

We use canonical correlation analysis (CCA) to estimate spatial filters and canonical components from training data. The use of CCA has been proven efficient in the past for decoding SSVEPs (Nakanishi et al., 2015) and ERPs (Spüler et al., 2014; Xu et al., 2018; Xiao et al., 2020). The approach presented here is derived from our previous work (Reichert et al., 2016, 2017) and closely related to the approach recently published (Reichert et al., 2020a). CCA successively determines coefficient vectors a and b that linearly combine two sets of variables X and Y such that the correlation of Xa and Yb is maximal:

$$(u, v) = \underset{a,b}{\operatorname{argmax}} \operatorname{corr}(Xa, Yb) \tag{1}$$

where u and v are the resulting canonical variables. In the present implementation, X represents the concatenation of EEG epochs and a serves as a spatial filter. To reveal the hemispheric differences that characterize the shift of attention to the left or right visual field, the difference wave following left target presentations and right target presentations is commonly computed. We model the difference wave by composing a

matrix Y that is a concatenation of identity matrices $I \in \mathbb{R}^{38\times38}$ weighted with the labels y_i, where $y_i I$ indicates that the participant paid attention to the green symbol and $-y_i I$ indicates that the participant paid attention to the red symbol. The kth column in the variable set Y represents the kth time point after stimulus onset in an epoch and can be considered a concatenation of impulse functions, i.e., a vector of zeros being at the kth sample 1 if the target was presented left and -1 if the target was presented right. With this matrix, the difference waves for each channel in X, composed of n epochs per hemifield of target presentation, can be easily calculated as $\hat{X} = n^{-1}Y^T X$. However, since we want to determine optimal spatial filters, we apply CCA to X and Y using the MATLAB® function *canoncorr* of the Statistics and Machine Learning Toolbox™. As a result, we reveal 14 vectors a, whose elements can be used as channel weights and 14 vectors b, whose elements depict the canonical difference waves. Since the canonical correlation decreases with each iteration of the CCA algorithm, we retain only vectors that achieve a significance level $p < 0.1$ according to the *canoncorr* function. This procedure is performed with an arbitrary training set of trials to estimate the weight vectors a and b needed to classify the attended symbol.

Decoding a Sequence of Attention Shifts

After we have determined a and b from training data, the target of a new sequence is detected as follows. We concatenate the epochs of the trial as X' and concatenate the corresponding weighted identity matrix $y_i I$ as Y', i.e., we initially assume that the green symbol was the target symbol. We then calculate the Pearson product-moment correlation ρ of $X'a$ and $Y'b$ for all vectors we retained after CCA and calculate the mean correlation $\bar{\rho}$. The stimulus sequences were designed such that the target randomly changed between the visual hemifields. Because there are only two alternatives, which were presented simultaneously on opposite sites, the sequence for red targets is the reverse of the sequence for green targets. Thus, if $\bar{\rho}>0$, the canonical difference waves of the EEG correlate with the canonical difference waves of the model functions corresponding to the sequence of green symbols, as assumed when modeling Y', and indicating that the participant intended to respond "yes." If $\bar{\rho} < 0$, the canonical difference waves of the EEG correlate with the canonical difference waves of the model functions corresponding to the sequence of red symbols, and we present "no" as feedback.

Evaluation of BCI Performance

During the experiment, we decoded all trials that followed the first two training blocks and presented the result as feedback. For this online decoding, we involved all available trials we had recorded by that time and did not discriminate between stimulus features. In contrast, we performed offline decoding by leave-one-out cross-validation (LOOCV) and determined decoding accuracies that can be achieved by varying stimulus features. To compare the outcomes using small training subsets with those achieved with larger training sets, we matched the sample sizes by random selection of trials from the larger training set. We repeated the random selection one hundred times and averaged the decoding accuracies achieved in the LOOCVs.

We performed ANOVA and paired Wilcoxon signed rank tests to evaluate the impact of stimulus features. To determine the chance level of the decoder empirically, we performed a permutation test. Specifically, we randomly permuted the labels "yes" and "no," which also implies randomized "target left" and "target right" assignments and performed LOOCV. This procedure was repeated 500 times. Afterward, we determined the 95% confidence interval from the distribution of decoding accuracies.

EOG Analysis

We pursued two strategies to evaluate a potential impact of eye movements on BCI performance. First, we applied our decoding approach to the EOG data that we recorded during the experiment and compared the accuracy achieved with that achieved with the parieto-occipital EEG. Second, we compared the EOG recorded during the experiment with the EOG calibration data that we recorded prior to the experiment. Therefore, we calculated the deflection of the hEOG as follows. The hEOG data were segmented according to the cue in the EOG calibration and each single stimulus in the BCI experiment, respectively. We involved a time interval of 750 ms length starting from the cue or stimulus onset and performed baseline correction according to the first 100 ms. Afterward, we selected the 25 highest absolute values across the interval, which corresponds to 100 ms of strongest hEOG deflection, and averaged these values.

RESULTS

Online BCI Performance

On average, 88.5% (σ = 7.8%) of participants' responses were correctly decoded with our decoding approach. Individual decoding accuracies ranged from 70.8% to 90.3%, while the chance level was 50%. The average accuracy corresponds to an information transfer rate of 3.02 bit/min, neglecting the time for asking questions and providing feedback. The decoding accuracy of questions with subjective answers (μ = 88.4%, σ = 8.4%) did not significantly differ from decoding accuracies of questions with obvious answers (p = 0.905; μ = 88.5%, and σ = 8.0%). Because we presented the same number of "yes" and "no" questions with objective answers in each run except in the last run, where answers were initially unknown, the sample sizes of the two classes were balanced, which reduces the probability that class sizes bias the classification. Consequently, the true positive rate of both classes was not statistically different (p = 0.298; "yes": μ = 89.4%, σ = 8.9%; "no": μ = 87.5%, and σ = 8.1%), indicating that the decoder was not biased. The reported decoding accuracies are summarized in **Figure 3**.

During the experiment, we retrained the classifier after each trial starting from the third run to provide the maximum number of trials available for model estimation. In practical use, the retraining would not be possible since we would not know the ground truth of the user's intention. Therefore, we estimated the accuracy that can be achieved with only two runs of training by repeatedly performing a LOOCV with matched sample size. On average, 85.3% (σ = 11.1%) of trials were correctly decoded, which

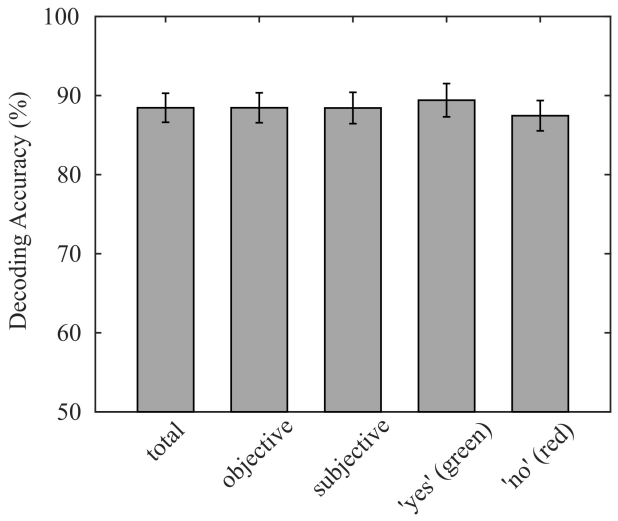


FIGURE 3 | Average decoding accuracies achieved online, validated on the entire data set, separated for objective and subjective questions, and separated for green and red targets. No significant differences were found for any sub group.

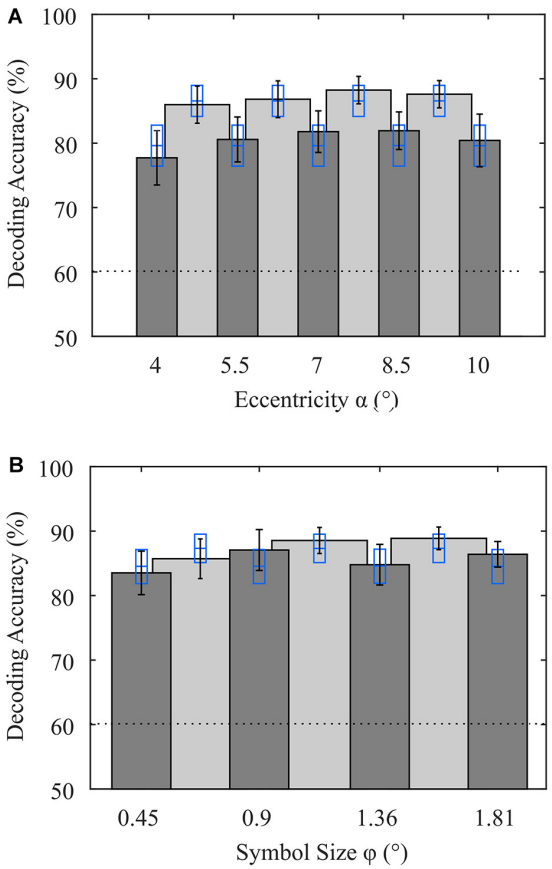


FIGURE 4 | Average decoding accuracies achieved with variations in stimulus features **(A)** eccentricity and **(B)** symbol size. Dark gray bars indicate that only trials with that parameter (eccentricity or symbol size) were included in the analysis, but the other parameter varied (symbol size or eccentricity). Light gray bars indicate that in the analysis, trials associated with the two adjacent visual angles were included (e.g., 4° and 5.5°). Error bars indicate the standard error of the mean. Blue rectangles indicate the mean and the standard error of accuracies achieved with all stimulus features involved but sample sizes of the training data sets matched. The dotted line denotes the upper bound of the guessing level, as determined using a permutation test.

demonstrates that the retraining of the classifier improved the overall performance of the BCI significantly ($p < 0.05$).

Evaluation of Stimulus Features

By using LOOCV, we maximize the amount of training data available for estimation of the spatial filters and the canonical difference waves required for detection of the attended symbol. We opted for LOOCV, because the number of samples available is small when validating subsets according to stimulus features. When all eccentricities and symbol sizes were included, as was the case for the online decoding, we achieved an average decoding accuracy of 88.6% (σ = 8.1%) using LOOCV. With this full data set, we performed a permutation test for each participant, which resulted in an upper threshold of 59.9% (σ = 0.8%) on average for the chance level. To investigate the impact of the stimulus' eccentricity, we performed LOOCV involving only trials where stimuli were presented at a specific visual angle, irrespective of symbol size. Likewise, we performed the same analysis for the stimulus feature symbol size. To prevent bias in the evaluation of the performance of a subset due to larger sample sizes in the training data, we matched the sample sizes as described in (2.6). As a result, we found that, on average over participants, the eccentricities $\alpha_1 = 4°$ and $\alpha_5 = 10°$ resulted in slightly lower decoding accuracies, but there was no significant difference between visual angles α (**Figure 4A**). For the symbol size feature, there was also no statistically significant difference (**Figure 4B**). Furthermore, to increase the sample size of the training data set, we grouped each visual angle with its adjacent visual angle. While the accuracies were generally higher with these larger data sets, presumably due to a better generalizable model estimation, decoding accuracies did not statistically differ between tested visual angles. The symbol size groups φ_2/φ_3 and φ_3/φ_4 achieved statistically significantly higher accuracies compared to φ_1/φ_2 ($p < 0.05$, uncorrected). When comparing the parameter subsets, involving all trials but with matched sample size, only $\alpha_3 = 7°$ and α_3/α_4 resulted in statistically significantly higher accuracies.

Finally, we also evaluated the performance of each parameter pair used in the experiment. Note that only 3 trials per run were available for each pair, resulting in 21 trials available for LOOCV. A 2-way ANOVA revealed no significant effect of the factors *eccentricity* (p = 0.98, $F_{4,143}$ = 0.1) or *symbol size* (p = 0.72, $F_{3,143}$ = 0.45). In **Table 1**, we show the results of the single parameter combinations. None of the combinations was significantly superior to another combination. Only the parameter combination (φ_3, α_3) achieved statistically significant higher decoding accuracy ($p < 0.05$, uncorrected) compared to classification independent from stimulus features with matched sample size. Each of the combinations achieved

TABLE 1 | Decoding accuracy achieved with single parameter pairs (standard deviation in parenthesis).

Visual angle	α_1	α_2	α_3	α_4	α_5
φ_1	77.7 (17.9) %	79.3 (16.9) %	–	–	–
φ_2	–	83.2 (13.0) %	82.5 (15.4) %	–	–
φ_3	–	–	79.6 (15.5) %	79.7 (14.6) %	–
φ_4	–	–	–	83.4 (17.6) %	80.4 (17.4) %

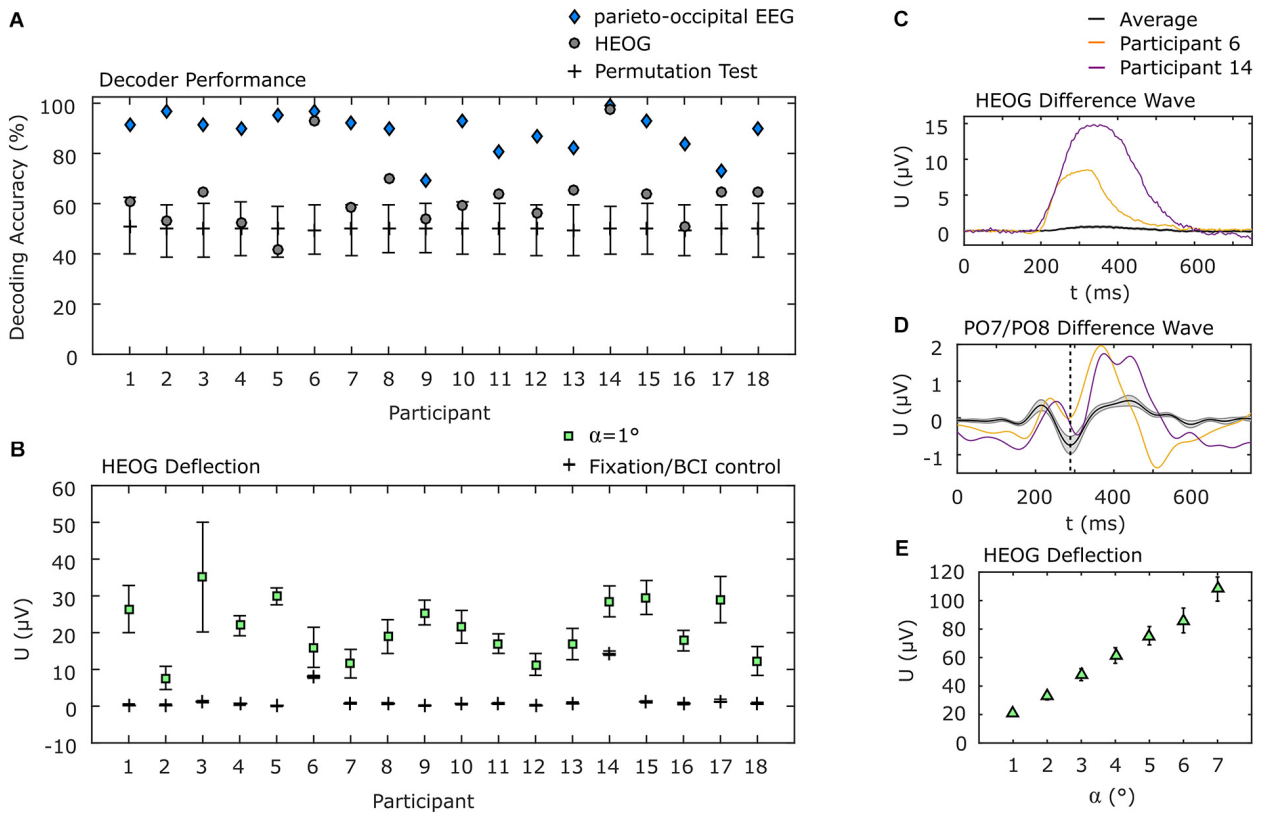


FIGURE 5 | EOG analysis. **(A)** Individual decoding accuracies achieved with EOG and with parieto-occipital EEG. EOG decoding accuracies for participants 6 and 14 were above 90% but below those achieved using EEG. Error bars indicate the 95% confidence interval of a permutation test. **(B)** Individual horizontal EOG deflections. Here the same two participants also show higher deflections during BCI use, but they are still below the deflection measured when 1° movements were requested in a calibration procedure. All other participants show almost no deflections during BCI use. **(C)** EOG difference wave between left and right target presentation. Participants 6 and 14 show considerably higher EOG deflection compared to the average. **(D)** PO7/PO8 difference wave. Participants 6 and 14 show the N2pc component (dashed vertical line), typically evoked during spatial attention shifts, similar to the average but with higher positive deflection afterward. **(E)** Deflection of the EOG as a function of the angle of gaze shift during EOG calibration. Error bars in **(B,E)** indicate the standard error of the mean.

individual maximum decoding accuracy in at least two participants. From the distribution of classification results independent from stimulus features with matched sample size, we repeatedly draw 8 decoding accuracies (in analogy to the 8 combinations) and determined the average of the maxima. This simulates the probability that a maximum value was achieved by an advantageous drawing of trials. We found no significant difference between maxima achieved with parameter combinations and maxima achieved by randomly drawing eight trial subsets from the entire data set. All these results indicate that the differences in stimulus features chosen in this study have no significant impact on the decoding accuracy achieved with the spatial attention paradigm.

Impact of Eye Movements

The use of only parieto-occipital electrodes reduces the probability that eye movements have a systematic impact on the decoding accuracy. However, we pursued two additional strategies to explore a potential impact of eye movements. Firstly, if eye movements nonetheless played a substantial role in discrimination of visual attention shifts, decoding the EOG data should result in higher accuracies than the EEG channels over brain areas attributed to visual processing. Therefore, we applied the same decoding approach to the EOG signals as we applied to the parieto-occipital EEG channels. The average decoding accuracy of 63.0% (σ = 13.6%) is significantly lower than that achieved using EEG data ($p < 0.001$). Notably, for

two of the participants, classification of EOG data resulted in an accuracy of above 90%, but the EEG accuracy was higher still for those participants (see **Figure 5A** for details). For the remaining participants, EOG decoding accuracy was within or slightly above the confidence interval for chance as determined by a permutation test.

In a second analysis, we determined the deflection of the hEOG during the experiment, according to the presentation side of the target, and compared it with the hEOG deflection obtained from defined eye movements. We found that most of the participants showed almost no hEOG deflection. Specifically, it was much lower than the smallest gaze angle of 1°, which we tested in the EOG calibration session. This finding is in concordance with the results obtained applying our decoding approach to EOG data only, leading to accuracies close to the guessing level for most of the participants. The two participants who showed high decoding accuracy based on EOG channels also showed highest hEOG deflections, but still below that of 1° gaze angle (**Figure 5B**). To further provide evidence that the BCI was not influenced by eye movements, even in the two participants showing higher EOG deflections compared to the remaining participants, we show the difference waves (ipsilateral target presentation subtracted from contralateral target presentations) of the hEOG (**Figure 5C**) and of the EEG signal at PO7/PO8 (**Figure 5D**) for these participants and compare it with the average signals from the remaining participants. While the hEOG was much larger in these two participants compared to the group average, the difference wave at PO7/PO8 shows the typical N2pc component around 288 ms, which is a marker for shifts in spatial attention. However, the eye movements also might propagate to these channels as indicated by the larger positive deflection. Finally, we show the hEOG deflections as a function of the gaze angle obtained in an EOG calibration session in **Figure 5E**. Comparison indicates that if participants had shifted their gaze directly to the target with the lowest eccentricity ($\alpha_1 = 4°$), an average hEOG deflection of 61.4 μV ($\sigma = 23.2$ μV) would be apparent. However, the average deflection during the BCI experiment was much lower (<1 μV for 13 participants, <1.5 μV for 3 participants, <8 μV for 1 participant and <15 μV for 1 participant).

DISCUSSION

The BCI implementation presented here, demonstrates that questions can reliably be answered with "yes" or "no" simply by directing visual spatial attention to one of two simultaneously presented colored symbols. Sensitivity to differences in stimulus features, specifically to the size and eccentricity of presented symbols, could not be found with statistical evidence. However, large symbols tended to lead to more accurate decoding, which suggests that even for persons with impaired vision, attention to a perceived color in the left or right visual hemifield might be sufficient to determine the shift of spatial attention for reliable communication.

Although the information transfer rate of binary classification is low by definition, communication on a "yes" or "no" basis could provide important assistance in maintaining social interaction for persons who cannot otherwise communicate. The fact that the BCI can be controlled independently of gaze shifts and is thus potentially accessible to severely disabled potential users may be deemed to compensate for the low information transfer. Gaze-independent BCIs with two answer options have indeed been implemented using several other modalities. For example, the covert shift of attention to auditory stimuli has been decoded using EEG-based BCIs, achieving a bit rate of 2.46 bit/min and an accuracy of 78.5% (Halder et al., 2010), which is below the performance achieved with our visual spatial attention approach. Another auditory approach achieved 4.98 bit/min neglecting inter-trial gaps and 85% and 77% accuracy, respectively, (Hill and Schölkopf, 2012; Hill et al., 2014). Using motor imagery of hand and foot movement to respond to auditorily presented questions, only two of ten healthy participants achieved effective control (Müller-Putz et al., 2013). In further studies, covert speech was performed in the form of mental repetition of the words "yes" and "no," which resulted in decoding accuracies of 63.2% (Sereshkeh et al., 2017a) and 69.3% (Sereshkeh et al., 2017b). An independent BCI based on SSVEPs and a non-spatial visual attention paradigm has produced an accuracy of 72.6% (Zhang et al., 2010). A new vibrotactile stimulation paradigm achieved an accuracy of 76.7% and a bit rate of 1.35 bit/min in comparison to the benchmark paradigm where accuracy was 65.6% and the bit rate was 0.61 bit/min (Jin et al., 2020). Vibrotactile stimuli were also tested for communication in six locked-in syndrome patients (Lugo et al., 2014), where the grand average accuracy was reported to reach 55.3%. In a follow-up study, an 86.7% decoding accuracy was achieved with vibrotactile stimulation and 83.3% using motor imagery in healthy controls, but only 63.1% accuracy was achieved with vibrotactile stimulation and 58.2% with motor imagery in a patient group (Guger et al., 2017). This dramatic reduction in decoding accuracy was seen in patients suffering from a motor neuron disease when the somatosensory and motor cortex were involved in the control strategy, but it remains unclear whether such a reduction in accuracy would also be expected using visual spatial attention to provide responses. In a case study investigating the ability of a completely locked-in patient to communicate "yes" and "no" by thinking the answer, over 70% accuracy was achieved using functional near infrared spectroscopy (Gallegos-Ayala et al., 2014). This approach was further investigated with healthy subjects where an accuracy of 75% was achieved (Hwang et al., 2016). While there is no firm evidence that it is possible to discriminate in EEG signals between simply thinking "yes" or "no," we have shown that the direction of visual spatial attention can be clearly discriminated, with a 88.5% decoding accuracy, on a binary basis in EEG recordings from healthy participants. Future studies are required to determine applicability in patient groups.

Importantly, the results of the current study suggest that the decoding of spatial attention shifts is largely independent of several empirically chosen parameters for stimulus presentation. The parameter choices were made according to commonly reported measures in N2pc-relevant literature (Luck and Hillyard, 1994b; Eimer, 1996; Luck et al., 1997; Hickey et al., 2009; Mazza et al., 2009; Grubert et al., 2017; Donohue et al.,

2018; Drisdelle and Jolicoeur, 2019). The independence of BCI performance from stimulus features is indicated by the high accuracy achieved online, where we trained and tested the BCI using all trial types irrespective of the symbol size and eccentricity. However, since the number of samples available to train a classifier was higher for the whole data set than that available for subsets that represent specific stimulus features, the trained model might have been estimated with better generalizability, leading to a higher accuracy. We therefore reduced the number of samples used to train the classifier on data including all possible stimulus features to match the sample sizes of the subsets based on particular stimulus features. None of the subsets of trials associated with single parameters for stimulus features led to significantly different accuracy when averaged across participants but we found a marginal significance for the eccentricity 7°. The combination of two subsets of adjacent parameter values increased the accuracy compared to single subsets but again, the major increase can be attributed to the greater number of samples. In a recent study, magnified and non-magnified symbols were presented at different eccentricities showing different amplitudes at different eccentricities, but no interaction was found between eccentricity and magnification (Papaioannou and Luck, 2020), i.e., the N2pc amplitude accompanying larger symbols did not differ from that observed when smaller symbols were employed. This lack of difference is in accord with our finding that alterations in symbol size did not result in a significant change in decoding accuracy, which makes the paradigm potentially suitable for persons who suffer from impaired vision, because the perception of the target color, as the to be attended feature, in the left or right visual field, independent of symbol shape, might be sufficient to decode the attention shift of those persons. Furthermore, visually impaired people might not be able to discriminate between different symbols in the same visual hemifield, which is why we did not present competing distractors, although they are assumed to increase relevant hemispherical differences in the EEG (Luck et al., 1997). In our experiment, participants achieved reliable control without presentation of competing distractors. Whether distractors could lead to a further increase in accuracy is one of the questions for future studies. Regarding eccentricity, there was trend in our data to better discrimination of spatial attention when targets were presented in the range of five to nine degrees visual angle. Papaioannou and Luck (2020) investigated the effect of eccentricities at visual angles smaller than 4° (which was the smallest in our study) and found that the N2pc amplitude was constant, even for stimuli near the midline, but the amplitude was significantly smaller at 8° visual angle. In contrast, the post-N2pc positivity was significantly larger at 8°. The difference in our findings could be explained by the fact that the algorithm we used automatically determined the relevant features from the EEG and thus, it is not clear whether the N2pc or the post-N2pc positivity is the main feature that discriminated between attention shifts with the different eccentricities. Further investigations are required in this regard. Also, the parameter space might be extended in future studies to determine individual boundaries at which the attentional shift is detectable from short sequences of stimuli. While decoding accuracy is reduced in visual P300-based BCIs with increasing eccentricity of the target (Treder and Blankertz, 2010) this study suggests that N2pc-based BCIs, which depend on shifted stimuli, are largely insensitive to eccentricity.

We evaluated the hEOG to exclude the possibility that eye movements have an impact on BCI control. For this purpose, we recorded defined saccades in a short session before the actual experiment. The hEOG amplitudes we recorded were in accordance with the findings of Lins et al. (1993). The hEOG activity we measured during BCI control was less in amplitude than measurements during execution of 1° saccades for all participants and close to zero for most of the participants. Only two participants unintentionally performed small saccades (below 1°) during BCI control, which might have biased the decoding accuracy, but their decoding accuracy using EEG was nonetheless greater than using EOG. However, since we involved only parieto-occipital electrodes showing the typical time course observed during visual spatial attention, it is unlikely that BCI control was achieved by eye movements.

The BCI we propose here is suited to communication of responses to yes/no questions simply by directing visual spatial attention to a colored, peripherally presented symbol in persons who are unable to move their eyes and has the potential to be used in the absence of high visual acuity. Our data suggest that the decoding accuracy of visual spatial attention is largely independent of symbol size and eccentricity. The new approach could potentially serve as an assistive communication technique for patients suffering from severe motor neuron diseases. Future work should involve evaluation of decoding accuracy in visually impaired individuals.

CONCLUSION

We implemented a BCI that decodes binary decisions from a short series of ERPs that solely reflect processes of spatial attention. We found that the symbol size and eccentricity of the bilaterally presented stimuli have a minimal impact on the overall accuracy of the BCI. Consequently, attention to simple features like color, independent of the stimulus' shape, might be sufficient to control such a BCI, rendering it promising for visually impaired end-users.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in the Open Science Repository of the Otto-von-Guericke University, Magdeburg (doi: 10.24352/UB.OVGU-2020-155) and in the BNCI Horizon 2020 web repository (http://bnci-horizon-2020.eu/database/data-sets).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the Otto-von-Guericke

University, Magdeburg. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CR and SD designed the experiment. CR implemented the BCI. CR and IFTC collected the data. CR, IFTC, and SD analyzed the data. CR, SD, CMS-R, H-JH, and HH interpreted the data. CR, SD, and CMS-R wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The work of this manuscript was partly funded by the German Ministry of Education and Research (BMBF) within the Research Campus STIMULATE under grant number "13GW0095D."

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2020.591777/full#supplementary-material

REFERENCES

Acqualagna, L., Treder, M. S., Schreuder, M., and Blankertz, B. (2010). "A novel brain-computer interface based on the rapid serial visual presentation paradigm," in *Proceedings of the 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC'10*, Buenos Aires, 2686–2689. doi: 10.1109/IEMBS.2010.5626548

Awni, H., Norton, J. J. S., Umunna, S., Federmeier, K. D., and Bretl, T. (2013). "Towards a brain computer interface based on the N2pc event-related potential," in *Proceedings of the International IEEE/EMBS Conference on Neural Engineering, NER*, San Diego, CA, 1021–1024. doi: 10.1109/NER.2013.6696110

Brouwer, A. M., and van Erp, J. B. F. (2010). A tactile P300 brain-computer interface. *Front. Neurosci.* 4:19. doi: 10.3389/fnins.2010.00019

Brunner, P., Joshi, S., Briskin, S., Wolpaw, J. R., Bischof, H., and Schalk, G. (2010). Does the "P300" speller depend on eye gaze? *J. Neural Eng.* 7:056013. doi: 10.1088/1741-2560/7/5/056013

Donohue, S. E., Bartsch, M. V., Heinze, H. J., Schoenfeld, M. A., and Hopf, J. M. (2018). Cortical mechanisms of prioritizing selection for rejection in visual search. *J. Neurosci.* 38, 4738–4748. doi: 10.1523/JNEUROSCI.2407-17.2018

Drisdelle, B. L., and Jolicoeur, P. (2019). Dissecting cognitive operations in difficult visual search using response-locked posterior contralateral negativity event-related potentials. *J. Cogn. Neurosci.* 32, 1–19. doi: 10.1162/jocn_a_01506

Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity. *Electroencephalogr. Clin. Neurophysiol.* 99, 225–234. doi: 10.1016/0013-4694(96)95711-9

Fahrenfort, J. J., Grubert, A., Olivers, C. N. L., and Eimer, M. (2017). Multivariate EEG analyses support high-resolution tracking of feature-based attentional selection. *Sci. Rep.* 7:1886. doi: 10.1038/s41598-017-01911-0

Farwell, L. A., and Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr. Clin. Neurophysiol.* 70, 510–523. doi: 10.1016/0013-4694(88)90149-6

Frenzel, S., Neubert, E., and Bandt, C. (2011). Two communication lines in a 3 × 3 matrix speller. *J. Neural Eng.* 8:036021. doi: 10.1088/1741-2560/8/3/036021

Gallegos-Ayala, G., Furdea, A., Takano, K., Ruf, C. A., Flor, H., and Birbaumer, N. (2014). Brain communication in a completely locked-in patient using bedside near-infrared spectroscopy. *Neurology* 82, 1930–1932. doi: 10.1212/WNL.0000000000000449

Grubert, A., Fahrenfort, J., Olivers, C. N. L., and Eimer, M. (2017). Rapid top-down control over template-guided attention shifts to multiple objects. *Neuroimage* 146, 843–858. doi: 10.1016/j.neuroimage.2016.08.039

Guger, C., Spataro, R., Allison, B. Z., Heilinger, A., Ortner, R., Cho, W., et al. (2017). Complete locked-in and locked-in patients: command following assessment and communication with vibro-tactile P300 and motor imagery brain-computer interface tools. *Front. Neurosci.* 11:251. doi: 10.3389/fnins.2017.00251

Halder, S., Käthner, I., and Kübler, A. (2016). Training leads to increased auditory brain-computer interface performance of end-users with motor impairments. *Clin. Neurophysiol.* 127, 1288–1296. doi: 10.1016/j.clinph.2015.08.007

Halder, S., Rea, M., Andreoni, R., Nijboer, F., Hammer, E. M., Kleih, S. C., et al. (2010). An auditory oddball brain-computer interface for binary choices. *Clin. Neurophysiol.* 121, 516–523. doi: 10.1016/j.clinph.2009.11.087

Heinze, H. J., Luck, S. J., Mangun, G. R., and Hillyard, S. A. (1990). Visual event-related potentials index focused attention within bilateral stimulus arrays. I. Evidence for early selection. *Electroencephalogr. Clin. Neurophysiol.* 75, 511–527. doi: 10.1016/0013-4694(90)90138-A

Hickey, C., Di Lollo, V., and McDonald, J. J. (2009). Electrophysiological indices of target and distractor processing in visual search. *J. Cogn. Neurosci.* 21, 760–775. doi: 10.1162/jocn.2009.21039

Hill, N. J., Ricci, E., Haider, S., McCane, L. M., Heckman, S., Wolpaw, J. R., et al. (2014). A practical, intuitive brain-computer interface for communicating "yes" or "no" by listening. *J. Neural Eng.* 11:035003. doi: 10.1088/1741-2560/11/3/035003

Hill, N. J., and Schölkopf, B. (2012). An online brain-computer interface based on shifting attention to concurrent streams of auditory stimuli. *J. Neural Eng.* 9:026011. doi: 10.1088/1741-2560/9/2/026011

Hopf, J. M., Boelmans, K., Schoenfeld, A. M., Heinze, H. J., and Luck, S. J. (2002). How does attention attenuate target-distractor interference in vision? Evidence from magnetoencephalographic recordings. *Cogn. Brain Res.* 15, 17–29. doi: 10.1016/S0926-6410(02)00213-6

Hwang, H.-J., Choi, H., Kim, J.-Y., Chang, W.-D., Kim, D.-W., Kim, K., et al. (2016). Toward more intuitive brain–computer interfacing: classification of binary covert intentions using functional near-infrared spectroscopy. *J. Biomed. Opt.* 21:091303. doi: 10.1117/1.jbo.21.9.091303

Jin, J., Chen, Z., Xu, R., Miao, Y., Wang, X., and Jung, T. P. (2020). Developing a novel tactile P300 brain-computer interface with a cheeks-stim paradigm. *IEEE Trans. Biomed. Eng.* 67, 2585–2593. doi: 10.1109/TBME.2020.2965178

Kaufmann, T., Schulz, S. M., Grünzinger, C., and Kübler, A. (2011). Flashing characters with famous faces improves ERP-based brain-computer interface performance. *J. Neural Eng.* 8:056016. doi: 10.1088/1741-2560/8/5/056016

Kelly, S. P., Lalor, E., Finucane, C., and Reilly, R. B. (2004). A comparison of covert and overt attention as a control option in a steady-state visual eyoked potential-based brain computer interface. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2004, 4725–4728. doi: 10.1109/iembs.2004.1404308

Kübler, A., Furdea, A., Halder, S., Hammer, E. M., Nijboer, F., and Kotchoubey, B. (2009). A brain-computer interface controlled auditory event-related potential (p300) spelling system for locked-in patients. *Ann. N. Y. Acad. Sci.* 1157, 90–100. doi: 10.1111/j.1749-6632.2008.04122.x

Lin, Z., Zhang, C., Wu, W., and Gao, X. (2007). Frequency recognition based on canonical correlation analysis for SSVEP-Based BCIs. *IEEE Trans. Biomed. Eng.* 54(6 Pt 2), 1172–1176. doi: 10.1109/TBME.2006.889197

Lin, Z., Zhang, C., Zeng, Y., Tong, L., and Yan, B. (2018). A novel P300 BCI speller based on the Triple RSVP paradigm. *Sci. Rep.* 8:3350. doi: 10.1038/s41598-018-21717-y

Lins, O. G., Picton, T. W., Berg, P., and Scherg, M. (1993). Ocular artifacts in EEG and event-related potentials I: scalp topography. *Brain Topogr.* 6, 51–63. doi: 10.1007/BF01234127

Luck, S. J., and Ford, M. A. (1998). On the role of selective attention in visual perception. *Proc. Natl. Acad. Sci. U.S.A.* 95, 825–830. doi: 10.1073/pnas.95.3.825

Luck, S. J., Girelli, M., McDermott, M. T., and Ford, M. A. (1997). Bridging the gap between monkey neurophysiology and human perception: an ambiguity

resolution theory of visual selective attention. *Cogn. Psychol.* 33, 64–87. doi: 10.1006/cogp.1997.0660

Luck, S. J., and Hillyard, S. A. (1994a). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology* 31, 291–308. doi: 10.1111/j.1469-8986.1994.tb02218.x

Luck, S. J., and Hillyard, S. A. (1994b). Spatial filtering during visual search: evidence from human electrophysiology. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 1000–1014. doi: 10.1037/0096-1523.20.5.1000

Lugo, Z. R., Rodriguez, J., Lechner, A., Ortner, R., Gantner, I. S., Laureys, S., et al. (2014). A vibrotactile P300-based brain-computer interface for consciousness detection and communication. *Clin. EEG Neurosci.* 45, 14–21. doi: 10.1177/1550059413505533

Matran-Fernandez, A., and Poli, R. (2017). Brain-computer interfaces for detection and localization of targets in aerial images. *IEEE Trans. Biomed. Eng.* 64, 959–969. doi: 10.1109/TBME.2016.2583200

Mazza, V., Turatto, M., and Caramazza, A. (2009). Attention selection, distractor suppression and N2pc. *Cortex* 45, 879–890. doi: 10.1016/j.cortex.2008.10.009

Müller-Putz, G. R., Pokorny, C., Klobassa, D. S., and Horki, P. (2013). A single-switch BCI based on passive and imagined movements: toward restoring communication in minimally conscious patients. *Int. J. Neural Syst.* 23:1250037. doi: 10.1142/S0129065712500372

Müller-Putz, G. R., Scherer, R., Brauneis, C., and Pfurtscheller, G. (2005). Steady-state visual evoked potential (SSVEP)-based communication: impact of harmonic frequency components. *J. Neural Eng.* 2:123. doi: 10.1088/1741-2560/2/4/008

Nakanishi, M., Wang, Y., Wang, Y. T., and Jung, T. P. (2015). A comparison study of canonical correlation analysis based methods for detecting steady-state visual evoked potentials. *PLoS One* 10:e0140703. doi: 10.1371/journal.pone.0140703

Papaioannou, O., and Luck, S. J. (2020). Effects of eccentricity on the attention-related N2pc component of the event-related potential waveform. *Psychophysiology* 57:e13532. doi: 10.1111/psyp.13532

Qu, J., Wang, F., Xia, Z., Yu, T., Xiao, J., Yu, Z., et al. (2018). A novel three-dimensional P300 speller based on stereo visual stimuli. *IEEE Trans. Human Machine Syst.* 34, 81–85. doi: 10.1109/THMS.2018.2799525

Raymond, J. E., Shapiro, K. L., and Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: an attentional blink? *J. Exp. Psychol. Hum. Percept. Perform.* 18, 849–860. doi: 10.1037//0096-1523.18.3.849

Reichert, C., Dürschmid, S., Bartsch, M. V., Hopf, J.-M., Heinze, H.-J., and Hinrichs, H. (2020a). Decoding the covert shift of spatial attention from electroencephalographic signals permits reliable control of a brain-computer interface. *J. Neural Eng.* 17:056012. doi: 10.1088/1741-2552/abb692

Reichert C., Tellez Ceja I. F., Dürschmid S. (2020b) Spatial attention shifts to colored items - an EEG-based Brain-computer interface. http://dx.doi.org/10.24352/UB.OVGU-2020-155

Reichert, C., Dürschmid, S., Heinze, H. J., and Hinrichs, H. (2017). A comparative study on the detection of covert attention in event-related EEG and MEG signals to control a BCI. *Front. Neurosci.* 11:575. doi: 10.3389/fnins.2017.00575

Reichert, C., Dürschmid, S., Kruse, R., and Hinrichs, H. (2016). An efficient decoder for the recognition of event-related potentials in high-density MEG recordings. *Computers* 5:5. doi: 10.3390/computers5020005

Salvaris, M., and Sepulveda, F. (2009). Visual modifications on the P300 speller BCI paradigm. *J. Neural Eng.* 6:046011. doi: 10.1088/1741-2560/6/4/046011

Sereshkeh, A. R., Trott, R., Bricout, A., and Chau, T. (2017a). EEG classification of covert speech using regularized neural networks. *IEEE ACM Trans. Audio Speech Lang. Process.* 25, 2292–2300. doi: 10.1109/TASLP.2017.2758164

Sereshkeh, A. R., Trott, R., Bricout, A., and Chau, T. (2017b). Online EEG classification of covert speech for brain-computer interfacing. *Int. J. Neural Syst.* 27:1750033. doi: 10.1142/S0129065717500332

Severens, M., Van der Waal, M., Farquhar, J., and Desain, P. (2014). Comparing tactile and visual gaze-independent brain-computer interfaces in patients with amyotrophic lateral sclerosis and healthy users. *Clin. Neurophysiol.* 125, 2297–2304. doi: 10.1016/j.clinph.2014.03.005

Spüler, M., Walter, A., Rosenstiel, W., and Bogdan, M. (2014). Spatial filtering based on canonical correlation analysis for classification of evoked or event-related potentials in EEG data. *IEEE Trans. Neural Syst. Rehabil. Eng.* 22, 1097–1103. doi: 10.1109/TNSRE.2013.2290870

Takano, K., Komatsu, T., Hata, N., Nakajima, Y., and Kansaku, K. (2009). Visual stimuli for the P300 brain-computer interface: a comparison of white/gray and green/blue flicker matrices. *Clin. Neurophysiol.* 120, 1562–1566. doi: 10.1016/j.clinph.2009.06.002

Tian, Y., Zhang, H., Li, P., and Li, Y. (2019). Multiple correlated component analysis for identifying the bilateral location of target in visual search tasks. *IEEE Access* 7, 98486–98494. doi: 10.1109/access.2019.2929545

Treder, M. S., Bahramisharif, A., Schmidt, N. M., Van Gerven, M. A., and Blankertz, B. (2011a). Brain-computer interfacing using modulations of alpha activity induced by covert shifts of attention. *J. Neuroeng. Rehabil.* 8:24. doi: 10.1186/1743-0003-8-24

Treder, M. S., and Blankertz, B. (2010). (C)overt attention and visual speller design in an ERP-based brain-computer interface. *Behav. Brain Funct.* 6:28. doi: 10.1186/1744-9081-6-28

Treder, M. S., Schmidt, N. M., and Blankertz, B. (2011b). Gaze-independent brain-computer interfaces based on covert attention and feature attention. *J. Neural Eng.* 8:066003. doi: 10.1088/1741-2560/8/6/066003

van Gerven, M., and Jensen, O. (2009). Attention modulations of posterior alpha as a control signal for two-dimensional brain-computer interfaces. *J. Neurosci. Methods.* 179, 78–84. doi: 10.1016/j.jneumeth.2009.01.016

Vialatte, F. B., Maurice, M., Dauwels, J., and Cichocki, A. (2010). Steady-state visually evoked potentials: focus on essential paradigms and future perspectives. *Prog. Neurobiol.* 90, 418–438. doi: 10.1016/j.pneurobio.2009.11.005

Walter, S., Quigley, C., Andersen, S. K., and Mueller, M. M. (2012). Effects of overt and covert attention on the steady-state visual evoked potential. *Neurosci. Lett.* 519, 37–41. doi: 10.1016/j.neulet.2012.05.011

Xiao, X., Xu, M., Jin, J., Wang, Y., Jung, T. P., and Ming, D. (2020). Discriminative canonical pattern matching for single-trial classification of ERP components. *IEEE Trans. Biomed. Eng.* 67, 2266–2275. doi: 10.1109/TBME.2019.2958641

Xu, M., Wang, Y., Nakanishi, M., Wang, Y. T., Qi, H., Jung, T. P., et al. (2016). Fast detection of covert visuospatial attention using hybrid N2pc and SSVEP features. *J. Neural Eng.* 13:066003. doi: 10.1088/1741-2560/13/6/066003

Xu, M., Xiao, X., Wang, Y., Qi, H., Jung, T. P., and Ming, D. (2018). A brain-computer interface based on miniature-event-related potentials induced by very small lateral visual stimuli. *IEEE Trans. Biomed. Eng.* 65, 1166–1175. doi: 10.1109/TBME.2018.2799661

Zhang, D., Maye, A., Gao, X., Hong, B., Engel, A. K., and Gao, S. (2010). An independent brain-computer interface using covert non-spatial visual selective attention. *J. Neural Eng.* 7:16010. doi: 10.1088/1741-2560/7/1/016010

Zhu, D., Bieger, J., Garcia Molina, G., and Aarts, R. M. (2010). A survey of stimulation methods used in SSVEP-based BCIs. *Comput. Intell. Neurosci.* 2010:702357. doi: 10.1155/2010/702357

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 01 December 2020
doi: 10.3389/fnins.2020.549524

WAUC: A Multi-Modal Database for Mental Workload Assessment Under Physical Activity

Isabela Albuquerque[1], Abhishek Tiwari[1], Mark Parent[1], Raymundo Cassani[1], Jean-François Gagnon[2], Daniel Lafond[2], Sébastien Tremblay[3] and Tiago H. Falk[1,4]*

[1] Institut National de la Recherche Scientifique - Énergie, Matériaux et Télécommunications, Université du Québec, Montréal, QC, Canada, [2] Thales Digital Solutions Inc., Québec City, QC, Canada, [3] École de Psychologie, Université Laval, Québec City, QC, Canada, [4] PERFORM Centre, Concordia University, Montréal, QC, Canada

Assessment of mental workload is crucial for applications that require sustained attention and where conditions such as mental fatigue and drowsiness must be avoided. Previous work that attempted to devise objective methods to model mental workload were mainly based on neurological or physiological data collected when the participants performed tasks that did not involve physical activity. While such models may be useful for scenarios that involve static operators, they may not apply in real-world situations where operators are performing tasks under varying levels of physical activity, such as those faced by first responders, firefighters, and police officers. Here, we describe WAUC, a multimodal database of mental Workload Assessment Under physical aCtivity. The study involved 48 participants who performed the NASA Revised Multi-Attribute Task Battery II under three different activity level conditions. Physical activity was manipulated by changing the speed of a stationary bike or a treadmill. During data collection, six neural and physiological modalities were recorded, namely: electroencephalography, electrocardiography, breathing rate, skin temperature, galvanic skin response, and blood volume pulse, in addition to 3-axis accelerometry. Moreover, participants were asked to answer the NASA Task Load Index questionnaire after each experimental section, as well as rate their physical fatigue level on the Borg fatigue scale. In order to bring our experimental setup closer to real-world situations, all signals were monitored using wearable, off-the-shelf devices. In this paper, we describe the adopted experimental protocol, as well as validate the subjective, neural, and physiological data collected. The WAUC database, including the raw data and features, subjective ratings, and scripts to reproduce the experiments reported herein will be made available at: http://musaelab.ca/resources/.

Keywords: mental workload, operator functional state, workload assessment, wearable sensors, multi-modal database, ambulant subjects

OPEN ACCESS

Edited by:
Ana Matran-Fernandez, University of Essex, United Kingdom

Reviewed by:
Christoph Tremmel, University of Essex, United Kingdom
Pietro Aricò, Sapienza University of Rome, Italy

****Correspondence:***
Isabela Albuquerque
isabela.albuquerque@emt.inrs.ca

Specialty section:
This article was submitted to Neural Technology, a section of the journal Frontiers in Neuroscience

Received: *06 April 2020*
Accepted: *08 September 2020*
Published: *01 December 2020*

Citation:
Albuquerque I, Tiwari A, Parent M, Cassani R, Gagnon J-F, Lafond D, Tremblay S and Falk TH (2020) WAUC: A Multi-Modal Database for Mental Workload Assessment Under Physical Activity. Front. Neurosci. 14:549524. doi: 10.3389/fnins.2020.549524

1. INTRODUCTION

The ability of humans to perform activities in an effective and sustainable way is crucial in situations where tasks are not fully automatic. In many scenarios, human performance might be safety-critical for human lives, such as in the case of tasks performed by aircraft pilots, firefighters, and first responders. In these cases, monitoring and quantifying the current capability of a subject

to correctly perform a task may be critical to prevent accidents and, consequently, save lives. In this context, the Operator Functional State (OFS) (Hockey, 2003a) research framework can be used to breakdown the relationship between human performance and the level of difficulty of the respective task (Ting et al., 2009). According to Hockey (2003b), OFS can be defined as "the variable capacity of the operator for effective task performance in response to task and environmental demands, and under the constraints imposed by cognitive and physiological processes that control and energize behavior." The operator functional state can be thus seen as the resulting interaction between a subject and a task, given specific environmental (e.g., noise, movement, and temperature) and psychophysiological (e.g., sleep loss, illness, fatigue, and anxiety) conditions. While the interplay between human, task, and environment influences task performance, as a compensatory mechanism attempts to sustain task performance levels, this change of resource allocation might increase psychophysiological activation (Ting et al., 2009). The capability of reliably monitoring OFS is key to constraining work shifts and adapting task demand levels, thus ensuring that operators are safely and optimally performing the designated tasks (Wilson and Russell, 2003a).

OFS is also defined as the processes that mediate task performance under stress and high workload (Hockey, 2003a). In this work, we are interested in the impact of mental workload (MW) on the operator functional state. Across several definitions in the literature, MW can be summarized as a construct that encompasses one's capability of performing a task along with the mental strain required for performing it under specific environmental conditions (Cain, 2007). The interest on studying this specific aspect that influences OFS stems from the importance of maintaining its balance during task execution. In case the operator needs to employ high levels of mental resources in order to achieve a required task performance for a long time, this might increase fatigue levels to such a point that the operator is no longer able to successfully perform the task. On the other hand, if the task is not demanding enough, it can lead to boredom and lack of engagement, which could also affect the operator's performance (Wilson and Russell, 2003a; Jasper et al., 2016). However, devising an objective strategy to assess MW is still an open challenge. One of the main reasons is its subjectivity, as different factors such as previous experience and temporal pressure might affect how each subject perceives the level of difficulty when performing a task (Charles and Nixon, 2019).

Mental workload can be assessed via subjective ratings, task performance outcomes, and psychophysiological measures. Each method considers different inputs and presents different time resolutions. Among those, strategies based on monitoring psychophysiological signals collected with wearable devices present the best temporal resolution, as they may capture OFS changes even before they are reflected in task performance (Ting et al., 2009). In the literature, clinical-grade devices are frequently employed to monitor psychophysiological responses (Yin and Zhang, 2017; Hefron et al., 2018). However, these devices usually require a long time to be setup, are not comfortable to be worn for extended periods, and might not allow the monitored subject to walk freely to perform their tasks. Thus, when considering real-world scenarios, where it is not possible to use clinical-grade devices to collect the required data, the use of wearable technology becomes key to enable MW monitoring. A further barrier to the wide deployment of MW models in real-world scenarios lies in the mismatch between training and testing conditions, as the former have typically relied on static subjects (usually sitting on a chair) performing tasks that do not demand intensive body movement. Representative examples include tasks such as the N-back (Milner, 1998) and the Cabin Air Management System (Sauer et al., 2000). As such, current models do not explicitly take into account factors such as movement artifacts and the interplay between physical activity and MW, and thus it is not possible to directly apply them in situations that involve ambulant subjects.

In order to decrease the gap between current research on MW assessment based on psychophysiological signals and real-world applications, here we describe a dataset collected using consumer-grade wearable devices in conditions that combine manipulation of MW levels with different levels of physical strain. The study involved 48 subjects and six neural and physiological modalities were acquired (electroencephalography, electrocardiography, breathing rate, skin temperature, galvanic skin response, and blood volume pulse), in addition to 3-axis accelerometry. Moreover, after each experimental session, subjective ratings of MW using the NASA Task Load Index questionnaire (Hart and Staveland, 1988) and physical fatigue using the Borg Scale (Borg, 1982) were collected.

We focus on providing resources for allowing the development of different strategies for assessing MW. By monitoring psychophysiological responses to tasks that gauge distinct levels of MW, it is possible to employ the collected signals to compute features that act as a proxy to quantify how much the OFS was affected by the respective change in the task. More specifically, we developed an experimental protocol using the Multi-Attribute Task Battery II (MATB-II) (Santiago-Espada et al., 2011) in which participants performed a cognitive task under two levels of MW (low, high) and under three levels of physical activity (no, medium, high) by either walking/running on a treadmill or riding a stationary bike. Recent works (e.g., Wilson and Russell, 2003b; Cassenti et al., 2010) have shown that the MATB-II better elicits MW than tasks typically reported in the literature, such as the N-back task (Milner, 1998), mental rotation (Johnson, 1990), and visual search (Shepard and Metzler, 1971). This experimental design allows investigating questions that remain elusive in the MW assessment literature, such as the interplay between different modalities, and the impact of increased physical activity and movement on MW correlates in terms of added artifacts, as well as what additional mental resources are drawn by the physical activity.

In the following, we summarize the main contributions of the WAUC dataset:

- Experimental setting more closely resembling real-world applications where mental and physical workload are simultaneously considered.
- Large number of participants (48) in comparison to similar studies.

- Two different physical activity modulators tested, namely, stationary bike and treadmill.
- Multiple signals modalities are provided, all time-synchronized during the collection process, to allow for multi-modal MW models to be developed.
- Ground-truth values for both mental and physical workload are provided, as well as perceived values measured via subjective ratings.

The remainder of this paper is organized as follows: in section 2, we provide a brief literature review on MW assessment and existing datasets. In section 3, we describe the experimental protocol. In sections 4, 5, respectively, we describe the experiments performed to validate the WAUC dataset and present the results. Conclusions are given in section 6.

2. RELATED WORK

In this section, we provide a brief overview of the literature related to the proposed dataset. We describe MW assessment methods based on subjective ratings, as well as methods that utilize neural and physiological data as source of information. Our dataset comprises multiple modalities collected from ambulant subjects, and to the best of our knowledge, no similar experimental setting was previously proposed in the MW assessment literature. Thus, due to lack of closely related work, we decided to highlight in this section previous work that utilized data collected when subjects were performing similar tasks to the ones considered in our experimental protocol. At last, we briefly describe similar datasets that provided multiple neural and physiological modalities but were not proposed with the aim of performing MW assessment.

2.1. Subjective Mental Workload Assessment

Given the importance of maintaining balanced levels of MW for successful and safe performance of critical tasks, several works in the literature proposed strategies for assessing this dimension of the OFS. Part of this previous work proposed to tackle the MW assessment problem using subjective measurements collected while the task was being performed. Such methods rely on participants periodically filling in a questionnaire with ratings related to their current OFS. Popular examples are the Subjective Workload Assessment Technique (SWAT) (Reid and Nygren, 1988), the NASA Task Load Index (NASA-TLX) (Hart and Staveland, 1988), and the Modified Cooper–Harper Scale (Wierwille and Casali, 1983). These methods feature a multi-scale grading of multiple MW aspects. One main drawback of using such questionnaires across multiple sessions, however, is that they do not take into account relative changes in the ratings for each time the subject answers the questions. In order to circumvent this issue, Vidullch et al. (1991) proposed the Subjective Workload Dominance (SWORD) technique, which consists of comparing pairs of tasks to build the so-called judgment matrix and then computing a final workload index. Similarly to the SWORD questionnaire, SWAT and NASA-TLX ratings can also be aggregated in order to provide a single workload measure.

Even though the methods based on subjective ratings collected at the same time the task is executed present a low-cost and easy-to-implement alternative to assess MW, this strategy presents critical limitations. As highlighted by Borghini et al. (2014), attending to a secondary rating task might increase the levels of working memory required to perform the main task. Thus, the sole act of filling the questionnaires may be responsible to changes in the reported MW. Moreover, those methods do not allow for continuous MW assessment and have poor temporal resolution. While reducing the intervals at which the operator needs to provide feedback could lead to improved temporal resolution, this may actually increase workload due to the number of interruptions to the task being performed.

2.2. Mental Workload Assessment From Neural and Physiological Data

Due to the aforementioned limitations of measuring subjective ratings in real-time during task performance, neural and physiological data collection and analysis have emerged as a promising alternative. Electroencephalography (EEG), for example, has been frequently used to monitor MW mostly due to its high temporal resolution in comparison to other neuroimaging techniques (Teplan et al., 2002). Recent work on EEG-based MW assessment has suggested that using hand-engineered features combined with a classifier to predict MW levels can achieve a satisfactory performance. Zhang et al. (2016) combined EEG spectral features with ensembles of Support Vector Machines to devise subject-specific workload models. Their proposed strategy achieved an average classification performance of 76.7% (5 classes) across seven subjects. The recent literature on EEG-based MW assessment has also been trying to leverage advances in representation learning methods powered by deep neural networks. Almogbel et al. (2018), for example, utilized convolutional neural networks to classify MW states on a task that simulates vehicle driving. Raw EEG was employed and the best model described in the paper obtained 95.3% accuracy on a binary classification task considering the single subject considered in the experimental protocol.

In addition to EEG, physiological responses related to heart rate changes and measured by electrocardiogram (ECG) have also been considered for MW modeling. Heart rate variability (HRV) is frequently employed as a correlate for MW based on cardiac activity. HRV has been shown to successfully capture changes in the sympathetic–parasympathetic balance and to be lowered by an increase in MW levels (Chaumet et al., 2019). In the context of controlling unmanned aerial vehicles, Jasper et al. (2016) verified whether HRV could be used as a predictor of operator MW in this scenario. Each one of the 20 participants simultaneously controlled multiple vehicles while their ECG was monitored. Paired t-tests between HRVs obtained in different parts of the experiment (e.g., planning and executing the task) confirmed the expected effect of lower HRV values as MW increased in terms of required vigilance and situational awareness. In addition to studying the relationship between HRV and MW, Castaldo

et al. (2017) also assessed its correlation with performance of repetitive tasks. Their study showed that eight HRV features, such as the mean of RR intervals and approximated entropy, presented a strong correlation with task performance (with $p > 0.05$), which suggests that HRV can also be used as a predictor of how successfully operators will execute a task.

Other studies in the operator functional state monitoring literature have attempted to leverage the complementary between different neural and physiological modalities to achieve improved MW assessment. In their study, Wilson and Russell (2003b) combined EEG, heart rate, eye movement, and respiration rate to model MW elicited using the MATB-II task. Features such as EEG power spectral density and ECG interbeat intervals were used as input to a neural network. The average achieved classification accuracy was 84.3% (high or low MW levels) with a training set, which simultaneously considered data from all subjects. Furthermore, Hogervorst et al. (2014) proposed to use subject-specific models to model MW for the N-back task based on EEG, ECG, skin conductance, respiration, and eye-related measures. Their findings, however, showed that the fusion of different modalities did not improve the performance on MW prediction in comparison to using individual signals.

2.3. Physical Activity During EEG Monitoring

The interest on employing EEG-based brain-computer interfaces to real-world applications where ambulant subjects are considered motivated a diverse body of work. Matthews et al. (2008) developed a low-power portable EEG monitoring device capable of long-term signal acquisition. Data were collected while subjects walked on a treadmill at a speed of 2 mph and performed mental tasks such as divide a number by 7 or played a first person video game. A performance of approximately 80% accuracy for binary MW (high or low) assessment was achieved. Snyder et al. (2015) aimed to isolate and investigate the effect of movement artifacts on EEG data. The proposed experimental protocol involved 10 subjects walking on a treadmill at four different speeds. Since the goal was to obtain pure gait-related artifacts, no mental task was performed during the experiment. Their analysis showed that independent component analysis yielded accurate localization for most of the artifacts components. Zink et al. (2016) studied the differences on brain activity due to movement and cognitive effort by proposing an experimental protocol that collected EEG while subjects were cycling on stationary bikes or freely biking. While biking, subjects were asked to perform a three-class oddball auditory task. EEG analysis showed a reduction in the P300 component in cases where subjects were performing physical activity on an unconstrained environment, suggesting that there exists an interplay between increase in cognitive load stemming from freely biking and perceived task difficulty.

2.4. Related Datasets

To the best of our knowledge, there is no publicly available multi-modal dataset for MW assessment based on wearables. In contrast, for orthogonal aspects of human cognitive states, such as emotion and affective states, there are a few popular multi-modal datasets with modalities similar to the ones collected here. As examples, we highlight the DEAP (Koelstra et al., 2011) and MAHNOB-HCI (Soleymani et al., 2011) databases. Both analyze human affective states and were recorded while subjects watched videos. In the case of DEAP, music video clips were used. For MAHNOB-HCI, in turn, videos clips were taken from different movies. In both datasets, modalities such as EEG, galvanic skin response (GSR), skin temperature, and breathing rate were made available and time-synchronized. In all cases, subjects were asked to remain still and seated while watching the video clips.

3. METHODS AND MATERIALS

3.1. Participants

As the experimental protocol involved sustained physical and mental strain for a considerable period of time, recruited subjects were submitted to a pre-screening process in order to prevent any potential risk during the data collection. Hence, candidates with cardiovascular diseases, neurological disorders, history of feeling dizzy, or fainting were not considered for the experiment. After the screening process, four participants were discarded and 48 were selected. Based on self-identified gender and the assigned physical activity modality (i.e., bike or treadmill) used during the experiment, a total of 22 participants used the treadmill (9 male, 13 female) and 26 performed the experiment using the bike (16 male, 10 female). The average age among the participants was 27.4 ± 6.6 years old. In order to avoid gender bias in our dataset, we intended to have a close number of male and female subjects, however, no candidate was rejected or accepted to participate in our experiment due to gender-related reasons. All participants consented to participating in the study and were remunerated (10 CAD/hour) for the time they spent at the experiment facility. The experimental protocol was approved by the Ethics Review Boards of INRS, Université Laval and the PERFORM Centre (Concordia University), the latter being the location in which data were collected.

Prior to arriving at the experiment facility, participants were advised to wear comfortable sportswear, and to not drink caffeinated beverages for at least 2 h prior to the beginning of the data collection. Before starting the task tutorial, participants were asked to read and sign (in case of agreement) a consent form containing a brief description of the goals of our project and allowing the use and sharing of the collected data for research purposes.

3.2. Experimental Protocol

The experimental protocol aimed at simultaneously modulating mental and physical workload. Participants executed mental tasks while performing physical activity. A full factorial (2 MW levels $\times$ 3 Physical strain levels) design was employed to capture main effects and interactions. The data collection protocol was preceded by a tutorial to make the participants familiar with the tasks. The tutorial consisted of slides presentation to explain the experimental procedure and the tasks to be executed. Subjects were allowed to take as much time as necessary to go through

the tutorial and to ask the experimenters as many questions as needed.

After ensuring the participant understood the tasks to be performed, the next step involved donning the devices. Subjects were first asked to wear a BioHarness 3 chest strap (Zephyr, USA) that integrates the ECG, breathing rate, and acceleration monitoring. Next, an Enobio portable 8-channel wireless EEG headset (Neuroelectrics, Spain) was placed. While electrode connections were checked and calibrated via the device's companion software, a second experimenter placed the E4 wristband (Empatica, USA) responsible to monitor skin temperature, GSR, and blood volume pulse (BVP).

To guarantee participants' safety during the experiment, a safety harness was placed at the participant's chest following the devices placement step mentioned above. This was only the case for the participants assigned to the treadmill task. For those assigned to the stationary bike, they were asked to adjust the seat according to their preference. In all cases, the height of the screen was adjusted lastly according to participants preferences. **Figures 1A,B** illustrate the experimental layout for the bike and treadmill, respectively, once all devices and safety features are in place. Before starting the data collection, each subject performed a practice session that corresponded to playing MATB-II for 10 min. While subjects were practicing, the experimenter observed whether they were capable of correctly performing each task.

Three levels of physical activity were considered: no movement, medium (treadmill: 3 km/h, bike: 50 rpm), and high movement (treadmill: 5 km/h, bike: 70 rpm). Since in the case of the stationary bike it was not possible to set the physical activity level for a fixed value during the experiment, we leveraged the training phase prior to each experimental section to let each participant get used to the speeds required during the data collection. Moreover, during each trial, the experimenter monitored whether the participant was deviating more than 5 rpm from the required speed and alerted the participant in case it did.

With respect to the MW levels elicited by MATB-II, two levels were considered, namely, low and high MW according to the task difficulty. In total, six possible combinations of joint MW and physical activity levels were tested. The experiment was then split into six sessions, each one corresponding to one of the six combinations previously described. The order in which each session was executed was counterbalanced among all the participants to avoid any ordering biases.

Before each session, data corresponding to two baseline periods were collected. During the first baseline, there were neither physical nor mental activity. Participants were asked to stand still and relax during 60 s. Following this relaxation period, the second baseline was recorded where the subject was asked to start moving according to the corresponding physical activity level assigned to the current session, but without at MW manipulation. Recordings of the second baseline period only began once the activity level reached a stable period and the recording then lasted for 2 min. Lastly, the experimenter gave the joystick to the participant and the 10-min session of combined mental physical effort started. After each task, a 5-min break was given. During this resting period, participants were asked to perform a subjective evaluation corresponding to the past task by filling the NASA-TLX questionnaire. They also reported their perceived fatigue level based on the Borg scale. Overall, the duration of each experimental session comprising the baselines, task, and subjective evaluation was 18 min, and the complete experimental protocol lasted roughly 2 h. **Figure 2** summarizes the entire experiment and shows the duration in minutes corresponding to each part of a complete session.

3.3. Stimuli

The MATB-II (Santiago-Espada et al., 2011) was employed for modulating the MW level on the participants. This set of tasks was originally devised to simulate different activities that need to be performed by an aircraft pilot. In our experiment, different mental strain levels are elicited by requiring the subjects to simultaneously perform three of the (four available) tasks involved in MATB-II, namely, system monitoring, tracking, and resource management. **Figure 3** shows a screenshot of the MATB-II interface, as seen by the participant. Note the top-right part of the screen was not used for the purposes of this study. An Xbox 360 controller was used to perform the three concurrent activities.

The system monitoring task (see top-left part of **Figure 3**) requires the participant to monitor four sliders and report deviations from their normal state. The two warning lights (seen as F5 and F6 in the figure) were not used in this study. In their normal states, sliders oscillate around the center position. In their deviation state, sliders start oscillating around the top or the bottom of the panel. Participants had to use the directional pad of the controller to report deviations (one direction was assigned to each slider). When reported, the concerned slider reverted to its normal state. In case the deviated sliders were not reported within 10 s, they were reverted to their normal state and a false alarm was recorded.

The tracking task (top-middle part of **Figure 3**), in turn, requires the participant to keep a target (a circular aim) within a square bounding box. As the trials progressed, the target started to move randomly. Participants had to use the joystick part of their controller to bring the target back near the center of the square. Lastly, the resource management task (bottom-center part of **Figure 3**) simulates the control of fuel reservoirs. Participants are asked to control pumps (which are subject to failure during the task) to transfer fuel across 6 reservoirs in order to keep the content levels of two main tanks (A and B) below a certain threshold. In particular, they were instructed to keep the level of the main tanks as close as possible to 2,500 units (this level is indicated by ticks on the sides of tanks A and B). However, fuel gradually depleted from tanks A and B. To keep the tanks at the aimed level, participants could use 8 pumps (labeled 1–8) to transfer fuel between the reservoirs. To activate pumps, participants had to use the second joystick of the controller to move the cursor and "click" on the pumps. When turned on, the pump would turn green. Pumps were configured to fail from time to time. When a pump failed, it turned red and was disabled. Pumps were automatically enabled for use after a while and the participant could resume using it if needed.

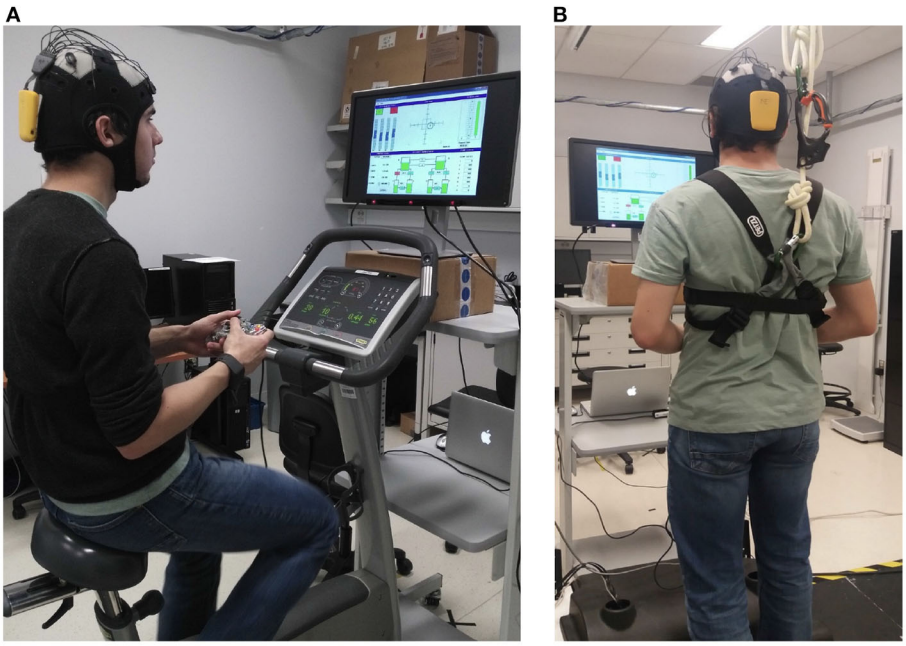


FIGURE 1 | Experimental set-up illustration for **(A)** bike and **(B)** treadmill sessions.

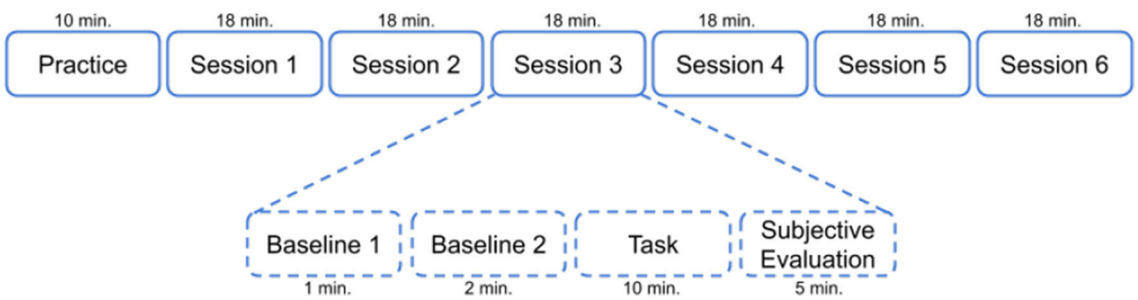


FIGURE 2 | Schematic of the steps executed by a participant during the experiment.

Modulation of the MW level relied on changing parameters in MATB-II. For example, for low MW cases, sliding bars speed, aim speed, volume of fuel in the reservoirs, and failure rate of the pumps were set to lower values. In the case of high MW, on the other hand, those parameters were set to larger values.

3.4. Subjective Evaluation

As mentioned previously, each experimental session within our protocol included a subjective evaluation. During this step, the NASA Task Load Index (NASA-TLX) questionnaire (Hart and Staveland, 1988) was employed. This set of questions was devised with the main purpose of providing a subjective metric for MW, which is less participant-specific and takes into account different factors resulting in mental strain.

The NASA-TLX questionnaire consists of the evaluation of six factors considered to impact MW, namely, mental demand, physical demand, temporal demand, performance, effort, and frustration. Subjects were asked to perform a self-evaluation of their mental/physical state with respect to each of these six dimensions using a 21-point scale.

In addition to the NASA-TLX questionnaire, we also employed the Borg fatigue scale (Borg, 1982) to assess the participant's fatigue level. They were asked to answer the following question using a scale from 6 to 20: "What physical effort and level of physical fatigue did the last segment impose on you?" We collected fatigue ratings before and after the 5-min break taken at the end of each experimental session.

3.5. Devices

In this study, three wireless wearable devices were employed to acquire data from 7 different modalities, as summarized in **Table 1**. The table also shows the sampling rate used during the data collection for each modality. The open-source software MuLES (Cassani et al., 2015) was utilized in order to allow

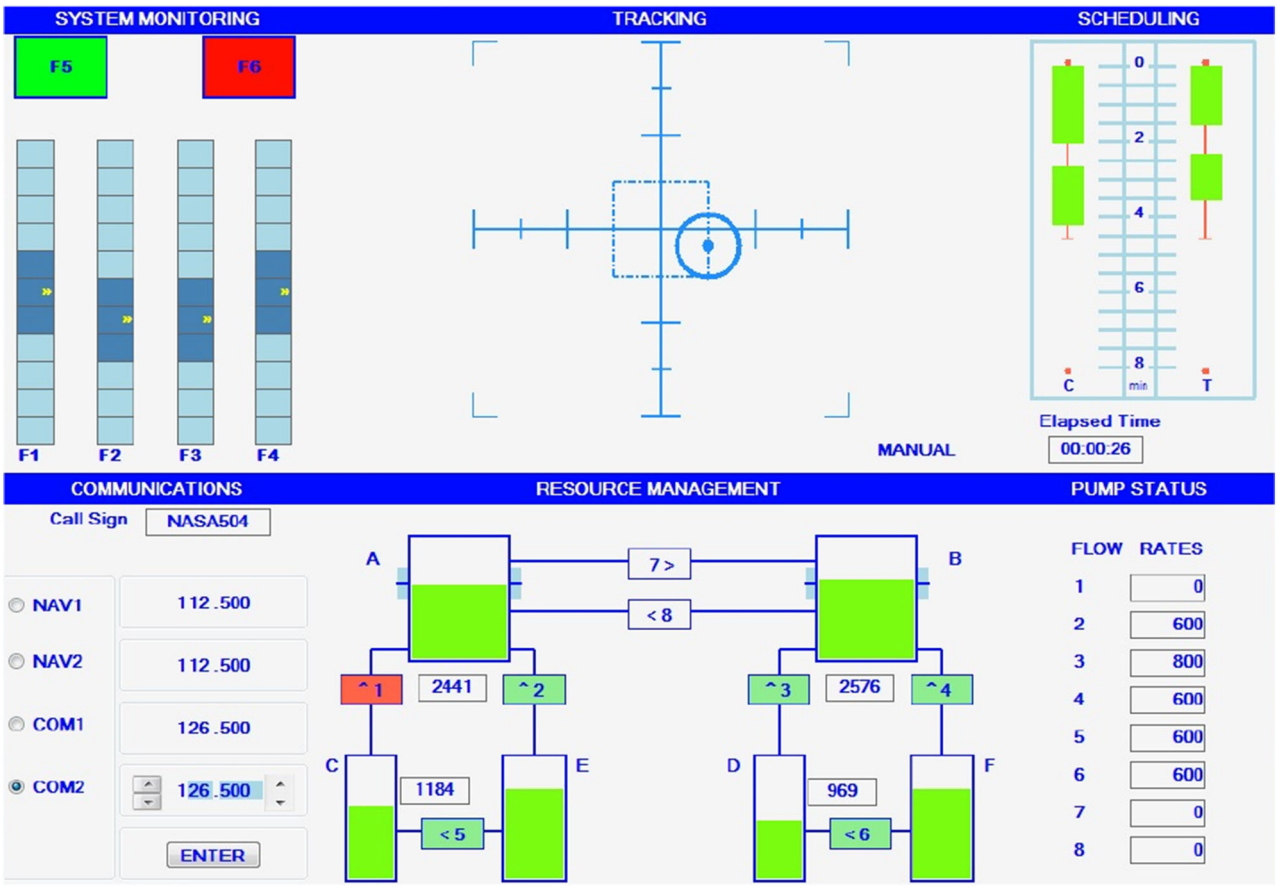


FIGURE 3 | Illustration of the Multi-Task Attribute Battery II (MATB-II) interface (figure obtained from NASA's website https://matb.larc.nasa.gov/).

TABLE 1 | Devices used in the data collection along with the respective acquired modalities and sampling rate.

	Modality	Sampling rate (Hz)
Enobio	EEG	500
Empatica E4	Skin temperature	4
	Galvanic skin response	4
	Blood volume pulse	64
	Acceleration	32
Bioharness3	ECG	250
	Breathing rate	25
	3-axis acceleration	18

simultaneous and synchronized acquisition of data streams from all devices. MuLES was also used to generate the synchronized markers indicating the beginning and the end of each phase of the experimental protocol. More details about each device is given below.

3.5.1. Enobio Headset

EEG data were collected using the 8-channel Neurolelectrics Enobio portable headset (Ruffini et al., 2007). The acquisition sampling rate was set to 500 Hz. Electrode positions according to the 10–20 system were P3, T9, AF7, FP1, FP2, AF8, T10, and P4. References were placed at Fpz and Nz. Since our study involved physical activity, we decided to use wet electrodes on the regions that would be likely affected by sweat during the experiments to avoid signal quality issues (Shu et al., 2019). Therefore, frontal and temporal regions were monitored using wet electrodes, while dry electrodes were used in the parietal region. **Figures 1A,B** illustrate Enobio's placement on the participant's head during the experiment.

3.5.2. E4 Wristband

The E4 wristband from Empatica was used to sample skin temperature, GSR, BVP, and acceleration at 4, 4, 64, and 32 Hz, respectively. The E4 was placed either on the left or right wrist, according to the participant's preference. In **Figure 1A**, it is possible to see the E4 positioned on the subject's right wrist. In the case of participants assigned to the treadmill, the E4 position was monitored during the experiment breaks in order to assure skin contact was not lost due to arm movements while running.

3.5.3. BioHarness3

The Bioharness3 acquired ECG, breathing rate, and acceleration at 250, 25, and 18 Hz, respectively. The device was supported by a chest belt containing two wet electrodes, one approximately placed at the tip of the sternum and another on the left side of the chest, both in direct contact with the skin. The length of the belt was carefully adjusted to avoid it from moving during the experiment. In **Figure 1B**, it is possible to observe the position of BioHarness3 belt across the subject's chest area. Note that this is for visual purposes only and in the actual experimental sessions, the belt was placed in direct contact with the participant's skin.

4. VALIDATION OF COLLECTED DATA

4.1. Validation Steps

In this section, we provide an overview of the analysis performed to validate the collected data, both in terms of subjective ratings and psychophysiological recordings. To validate the data obtained from the subjective evaluations, a mixed model analysis of variance (ANOVA) was used for each NASA-TLX dimension and the Borg scale values. As this experiment aimed to test the effect of different experimental conditions on the collected subjective ratings, a repeated measures design was used in order to take into account the within-subject variability on the data. For each aspect considered in the subjective evaluation, MW (with low or high levels) and physical workload (with no, medium, high intensity) were considered as within-subject independent variables, whereas equipment (bike or treadmill) was considered as the between-subject independent variable.

In addition to the ANOVA, we empirically analyzed the changes on the distribution of NASA-TLX dimensions ratings for each different physical strain level. With this analysis, a visual depiction of how different physical activity levels impact the subjective perception of different NASA-TLX factors can be seen. To this end, each rating was first mapped to a binary value (low or high), considering as threshold the respective average rating calculated per subject taking into account all experimental sessions. We then presented for each physical workload level the total number of sessions rated as high for low/high MW sessions. Moreover, we performed the same analysis considering the subjects grouped according to the equipment to manipulate physical strain.

Validation then proceeded by attempting to perform binary classification of MW levels using features commonly reported in the literature and exploring the changes in performance resultant from varying physical workload conditions. It is important to emphasize that as the goal of this paper is to describe the new dataset and validate its use for the purpose intended, achieving state-of-the-art MW level prediction performance is not a priority and exploring the use of new features and/or classifiers is left for future work.

In the following subsection, the features used for benchmark MW classification are described.

4.2. Features

For EEG data, signals were downsampled to 250 Hz and bandpass filtered with a bandwidth 1–45 Hz. Wavelet-enhanced Independent Component Analysis (wICA) (Castellanos and Makarov, 2006) was used to reduce the impact ocular and muscular artifacts as it has shown reliable performance on MW assessment across different groups of features (Albuquerque et al., 2019). As the multi-task nature of MATB-II requires frequent changes in gaze position during the experiment, using an enhancement method capable of removing eye-related artifacts is of great importance. Features were then computed from the wICA-enhanced signal over 4-s long epochs with no overlap between consecutive windows. For classifying mental and physical workload levels, classical spectral features were considered, namely power spectral density (PSD) at delta (1–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (30–45 Hz) frequency sub-bands.

In the case of the physiological modalities collected using the Empatica E4, features were computed over 30-s windows with no overlap between consecutive windows. Mean, median, standard deviation, maximum, and minimum values over the 30-s window were considered as features for classification. In the case of skin temperature, acquired signals were pre-processed to remove high-amplitude peaks artifacts.

For the ECG signal, in turn, a bandpass filter was performed between 5 and 25 Hz to enhance the QRS complex peaks. Visual analysis was then used to remove segments with no clear RR intervals. This was followed by an energy-based QRS detection algorithm (Behar et al., 2014), which is an adaption of the popular Pan & Tompkins algorithm (Pan and Tompkins, 1985). The RR series obtained was further filtered to remove outliers using range-based detection ($\geq$ 280 and $\leq$ 1,500 ms), moving average outlier detection, and a filter based on percent change in consecutive RR values ($\leq$ 20%), as implemented in Behar et al. (2018). Finally, benchmark time- and frequency-domain heart rate variability (HRV) features were extracted from each session using 5-min windows with a 4-min overlap. The HRV feature set and the window size selection was done based on recommendations made in Camm et al. (1996). The time domain features included mean, standard deviation, and coefficient of variation, while the frequency domain features were high frequency power (HF), normalized HF, low frequency power (LF), normalized LF, very low frequency power, and the ratio between HF/LF.

For the breathing signal, downsampling was first performed from 18 to 6 Hz. A low-pass filter was then applied to remove noise (Chebychev, 2 Hz, 8th order). Following this, descriptive statistical features that include, average, median, standard deviation, minimum, maximum, delta, range, coefficient of kurtosis, and skewness of the signal were calculated. Further, breathing spectrum is sometimes divided into four equally spaced bands between 0 and 0.4 Hz. To explore influence of higher frequency, the spectrum was divided into 5 equally spaced bands between 0 and 1 Hz and power in each of the bands was used as a spectral breathing feature.

5. VALIDATION RESULTS AND DISCUSSION

5.1. Subjective Ratings Analysis

5.1.1. Repeated Measures ANOVA

Table 2 reports the results for multiple mixed model ANOVA[1] performed on the subjective ratings in terms of partial effect size (n_p^2) and p-value. We observe that all evaluated subjective ratings were significantly ($p < 0.001$) affected by changes in the type of equipment used to modulate physical strain levels. Similarly, a significant effect (with $p < 0.05$) of MW (represented as MW in the table) manipulation was found for all subjective ratings. Physical workload (PW), in turn, was found to significantly affect (with $p < 0.05$) all subjective ratings except *Performance*. By observing the descriptive statistics of this factor in **Table A1** in the Appendix, it can be seen that for all physical activity levels, the average of this NASA-TLX dimension was approximately equal to 15 and 12 for low and high MW sessions, respectively. Interestingly, mental demand ratings are significantly changed by manipulation on physical strain, which indicates that there might be an interplay between physical activity and perceived MW, further confirming the importance of collecting the proposed dataset. No significant interactions were found between MW and equipment, as well as between physical activity and equipment. Finally, no interactions between MW and physical activity were found for all subjective measurements except *Effort*.

5.1.2. Distribution of Binary TLX Dimensions

Figure 4 shows the percentage of "high" ratings for each NASA-TLX dimension, considering low and high MW sessions separately (represented in blue and orange, respectively). In order to inspect the effect of changes in physical activity, each radar chart accounts only for data collected under a single physical activity level. Intuitively, we expect high MW sessions to present a higher number of "high" ratings for some of the TLX dimensions such as mental demand. On the other hand, in the case of performance, we suppose a lower number of "high" ratings will be obtained for high MW sessions.

Overall, when comparing the radar charts for data obtained under different physical activity conditions, we notice that the number of high-rated sessions for mental demand increased. Thus, participants found high physical workload sessions more demanding than sessions where there was no physical activity to be performed. We believe this aspect further indicates that mental and physical workload are confounded and this particular relationship should be closely investigated by future research using the described dataset. Moreover, it is possible to observe that the number of "high" ratings for factor *Performance* has not drastically changed when physical workload increased. As we previously highlighted, this was similarly observed in the results obtained by the ANOVA study presented in **Table 2**. We believe this indicates that, as described by the OFS framework, participants need to increase their physical and mental demand in order to maintain a certain overall performance level.

Moreover, as subjects performed physical activity using either a treadmill ($n = 22$) or a stationary bike ($n = 26$), radar charts for binary TLX ratings are also computed based on the equipment. Different patterns are expected based on equipment used, as for example, participants on the treadmill were holding the controller, thus could not use their arms to help with balance, which could induce changes in cognitive load. Radar charts obtained with treadmill and bike data are shown in **Figures 5**, **6**, respectively. Overall, distributions are found to be indeed different for most of the dimensions/experimental conditions for both equipment. More specifically, by comparing **Figures 5A**, **6A**, it can be seen that for sessions where no physical activity was required and a low MW task was performed, a higher percentage of subjects rated the *Mental Demand* dimension as high for the treadmill case. Despite the fact most of subjects rated this sessions as low physical demand, it is believed that this indicates that as subjects were standing during these sessions, this "extra" physical strain (in comparison to the bike) might be the responsible for increasing the perceived mental demand. Interestingly, in the case high MW sessions performed using a stationary bike, a higher percentage of subjects rated the *Effort* dimension as high.

5.2. Classification of MATB-II Mental Workload Levels

Binary MW classification (low vs. high values) was explored using the MATB-II difficulty level as the ground truth. We consider three different cross-validation strategies to train and evaluate classifiers: **(i) Mixed-subjects:** we pool the data from all subjects and use a five-fold cross-validation scheme to split it. This process is repeated 50 times after shuffling the dataset to obtain different cross-validation folds. We report the average performance across the 50 repetitions. **(ii) Intra-subject:** We train one classifier per subject using five-fold cross-validation and report the average performance across all subjects. **(iii) Leave-one-subject-out:** Classifiers are trained with data from all but one subject and then evaluated on data from the subject left out. In this case, we report the average performance across the models obtained when each subject was left out of training. In all cases, Random Forest classifiers with 15 estimators were employed and the performance achieved in terms of the area under the receiving operator curve (AUC) is reported in terms of the average and standard deviation. Notice that it was not possible to apply intra-subject and leave-one-subject-out cross-validation schemes on models trained with ECG and breathing features because the number of data points per subject after feature extraction was considerably low (only two examples per experimental session).

Table 3 presents the classification results obtained using features computed from each modality individually for the no, medium, and high physical workload conditions, as well as for all conditions combined. Overall, we observe that EEG and breathing features presented the best average performance

[1] Although the residuals from the obtained model are not Gaussian according to the Kolmorov–Smirnov test (with significance level of 95%), we observed that the histograms of residuals present a shape similar to a Gaussian distribution for all the considered subjective measures, except Frustration.

TABLE 2 | Partial effect size (η_p^2) obtained from repeated measures analysis of variance (ANOVA) for subjective ratings (MW, mental workload; PW, physical workload).

		Equipment	MW	MW × Equipment	PW	PW × Equipment	MW × PW
NASA-TLX	Mental demand	0.897*	0.555*	0.002	0.219*	0.006	0.026
	Physical demand	0.857*	0.231*	0.003	0.723*	0.055	0.014
	Temporal demand	0.866*	0.602*	< 0.001	0.350*	0.022	0.002
	Performance	0.952*	0.679*	0.015	0.062	0.005	0.042
	Effort	0.909*	0.593*	0.013	0.376*	0.031	0.066†
	Frustration	0.739*	0.445*	0.022	0.097†	0.008	0.041
Borg scale	Before break	0.967*	0.437*	0.006	0.719*	0.056	0.006
	After break	0.961*	0.174†	0.059	0.619*	0.062	0.038

*p-value ≤ 0.001, †0.001 < p-value ≤ 0.05, NO SYMBOL: p-value > 0.05.

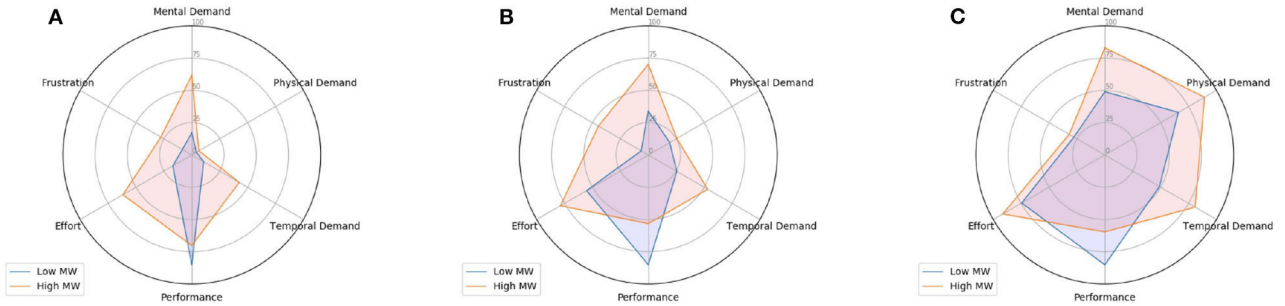


FIGURE 4 | Percentage of high-rated TLX dimensions (using the average value as threshold) per physical activity level. In this case, subjects that performed physical activity using both bike and treadmill are considered. **(A)** No physical activity. **(B)** Medium physical activity. **(C)** High physical activity.

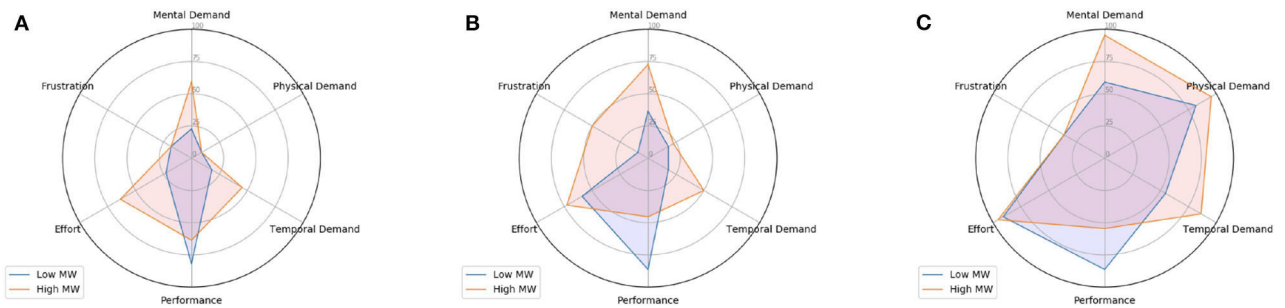


FIGURE 5 | Percentage of high-rated TLX dimensions (using the average value as threshold) per physical activity level. In this case, subjects that performed physical activity using **only the treadmill** are considered. **(A)** No physical activity. **(B)** Medium physical activity. **(C)** High physical activity.

and lowest standard deviation while BVP presented the lowest average performance. Interestingly, classifiers trained on all the conditions combined resulted in the lowest performance, suggesting that a hierarchical classification scheme may be needed where physical workload is first estimated and a PW-specific MW classifier is used. These findings corroborate the hypothesis of an interplay between mental and physical workload.

As expected, we observe that individualized models (intra-subject cross-validation) yielded the best performance across all considered modalities. However, this approach requires collecting data and training an MW classifier for each new subject to be monitored, which makes it less practical for real-world applications scenarios. From this perspective, leave-one-subject-out cross-validation appears to be the best strategy to be adopted, since no calibration is required prior to using the obtained MW classifier on new subjects. On the other hand, the overall poor performance of the obtained classifiers under this cross-validation scheme as presented in **Table 3** indicates that even though the considered features showed predictive power for MW for individual subjects, they are not robust to cross-subject variability.

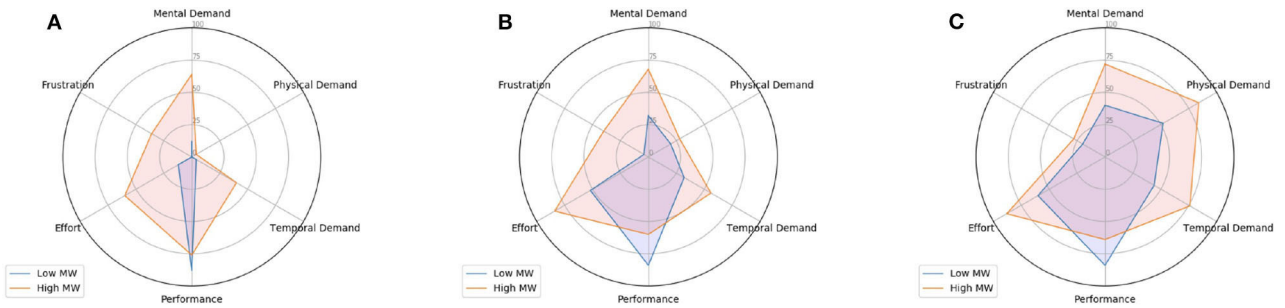


FIGURE 6 | Percentage of high-rated TLX dimensions (using the average value as threshold) per physical activity level. In this case, subjects that performed physical activity using **only the bike** are considered. **(A)** No physical activity. **(B)** Medium physical activity. **(C)** High physical activity.

TABLE 3 | Mean and standard deviation of area under the receiving operator curve (AUC) values obtained for binary mental workload classification when considering a model trained with data from all subjects, one model per subject and leave-one-subject-out validation.

Modality	Condition	AUC—Mixed subjects	AUC—Intra-subject	AUC—Leave-one-subject-out
EEG	No	0.774 ± 0.008	0.823 ± 0.139	0.523 ± 0.073
	Med	0.936 ± 0.004	0.927 ± 0.110	0.511 ± 0.093
	High	0.945 ± 0.004	0.929 ± 0.099	0.518 ± 0.112
	All	0.868 ± 0.004	0.805 ± 0.147	0.500 ± 0.049
Temperature	No	0.679 ± 0.026	0.846 ± 0.258	0.514 ± 0.142
	Med	0.641 ± 0.028	0.830 ± 0.279	0.509 ± 0.125
	High	0.656 ± 0.026	0.787 ± 0.303	0.506 ± 0.122
	All	0.594 ± 0.016	0.632 ± 0.282	0.514 ± 0.069
GSR	No	0.712 ± 0.025	0.882 ± 0.233	0.498 ± 0.144
	Med	0.761 ± 0.027	0.923 ± 0.169	0.522 ± 0.159
	High	0.692 ± 0.026	0.827 ± 0.256	0.557 ± 0.135
	All	0.661 ± 0.015	0.711 ± 0.264	0.519 ± 0.068
BVP	No	0.580 ± 0.029	0.720 ± 0.255	0.512 ± 0.109
	Med	0.624 ± 0.029	0.751 ± 0.258	0.539 ± 0.078
	High	0.584 ± 0.028	0.744 ± 0.249	0.494 ± 0.098
	All	0.562 ± 0.016	0.644 ± 0.183	0.481 ± 0.065
ECG	No	0.778 ± 0.016	-	-
	Med	0.780 ± 0.018		
	High	0.753 ± 0.026		
	All	0.748 ± 0.011		
Breathing	No	0.913 ± 0.011	-	-
	Med	0.892 ± 0.013		
	High	0.903 ± 0.012		
	All	0.865 ± 0.011		

In comparison to previous work that also considered MATB-II to modulate mental strain, we observe that the results presented in **Table 3** are in-line with the performances previously reported in the literature for experimental setting that considered static subjects. Specifically, Wilson and Russell (2003b) obtained 87.2% using only EEG spectral features, while we were able to achieve an average accuracy of 86.8% when taking into account all the physical workload levels altogether and a model trained using mixed subjects cross-validation (as in Wilson and Russell, 2003b). Similarly to our results, Wilson and Russell (2003b) also observed a decrease in the classification performance when only features computed from physiological modalities were considered. Note that although we obtained similar findings, the study in Wilson and Russell (2003b) only involved seven participants, as opposed to 48 in our case, and different approaches were considered to extract features and design classifiers, rendering the reported performance not directly comparable with results presented herein.

When observing the effect of increasing physical workload on the classification results, it can be seen that, in the case of mixed-subjects and intra-subject cross-validation schemes, EEG-based models obtained better performance when physical strain increased. This might be caused by an increase in the actual perceived MW during the task due to the extra effort not only in performing the physical activity, but also, for example, the increased mental resources used to avoid falling from the treadmill. This, added to the findings presented by Zink et al. (2016), which observed a decrease in the P300 component of EEG data in case subjects were biking in an outdoor environment, provides further evidence of the existence of an interplay between physical activity and perceived MW. For the other physiological features, in turn, the best classification performance was usually achieved in the no/medium PW condition. As the literature on movement artifact removal is more scarce for physiological signals, the findings in **Table 3** suggest that new enhancement algorithms may be needed, particularly for the high PW conditions.

5.3. Comparing Classification Performance: Bike vs. Treadmill

Recent research has shown that a human's attention to targets is reduced when walking relative to when standing still, due to processing demands produced by visual and inertial stimulation (Ladouce et al., 2019). As such, varying MW prediction capability is hypothesized based on the physical activity equipment used. **Table 4** shows the resulting AUC values for binary MW classification when using the treadmill or the stationary bike, as well as with both conditions combined. As can be seen, for all modalities, except ECG and breathing, average AUC values were higher in the treadmill condition. For EEG, these findings corroborate those of Ladouce et al. (2019).

TABLE 4 | Mean and standard deviation of area under the receiving operator curve (AUC) values obtained for binary mental workload classification under different signal modalities and physical activity equipment.

Modality	Equipment	AUC
EEG	Treadmill	0.924 ± 0.005
	Bike	0.801 ± 0.007
	All	0.868 ± 0.004
Temperature	Treadmill	0.629 ± 0.022
	Bike	0.626 ± 0.023
	All	0.594 ± 0.016
GSR	Treadmill	0.735 ± 0.022
	Bike	0.666 ± 0.020
	All	0.661 ± 0.015
ECG	Bike	0.534 ± 0.024
	All	0.562 ± 0.016
	Treadmill	0.762 ± 0.017
	Bike	0.773 ± 0.013
	All	0.748 ± 0.011
Breathing	Treadmill	0.875 ± 0.012
	Bike	0.876 ± 0.013
	All	0.865 ± 0.011

TABLE 5 | Mean and standard deviation of area under the receiving operator curve (AUC) values obtained for binary mental workload classification simultaneously considering EEG, skin temperature, GSR, and BVP features.

	AUC—Cross-subject	AUC—Leave-one-subject out
No	0.993 ± 0.006	0.561 ± 0.159
Med	0.998 ± 0.001	0.540 ± 0.253
High	0.998 ± 0.002	0.542 ± 0.217
All	0.995 ± 0.003	0.463 ± 0.115

5.4. Multi-Modal Mental Workload Classification

Lastly, we investigate whether performing MW classification on features computed from different modalities improves the obtained performance. For that, we consider feature-level fusion of EEG, skin temperature, GSR, and BVP features. To synchronize the features between modalities collected with different sampling rates, we average consecutive data points in order to obtain a single data point for each window of 60 s. This process resulted in a total of 10 examples per experimental session, each containing 47 features (32 EEG + 15 from the peripheral signals). In **Table 5**, we present the resulting AUC for models trained using mixed subjects and leave-one-subject-out cross-validation strategies. Note that we did not include ECG and breathing rate features as this would result in too few data points per subject. Moreover, we did not consider inter-subject cross-validation in this experiment for similar reasons.

When comparing the results presented in **Table 3** and **Table 5**, we observe that considering features from multiple modality provided an improvement in the classification performance in almost all the considered cases. Interestingly, we observe that in the case of mixed subjects cross-validation, the multi-modal approach presented improved robustness to an increase physical workload levels, indicating that the simultaneous use of multiple modalities to perform MW assessment might be key to design reliable systems.

5.5. Future Research Directions

We believe the WAUC dataset will enable research on several aspects of mobile brain–machine interfaces for practical everyday settings. The following list summarizes the main topics and problems that can be explored within further in-depth analysis of the WAUC dataset:

- Investigate the interplay between physical activity and MW on neural and physiological responses.
- Study the impact of physical strain on the interplay between increased levels of expertise on performing MATB-II and perceived MW (Borghini et al., 2017).
- Develop EEG artifact removal strategies that specifically address noise generated by physical activity for signals collected with low-density devices.
- Devise methods to detect variations on the intensity of MW instead of classifying a specific level.

- Leverage recent developments of deep neural networks to learn representations, which are invariant to subject-specific information in order to improve the performance under real-world scenarios where data from a new subject are not available at training time.
- Develop features tailored to improve robustness to movement artifacts and cross-subject variability.
- Devise models that are able to adapt to changes in the signal-to-noise ratio, as well as to new subjects.

6. CONCLUSIONS

Operator function state monitoring is critical for optimizing human resources allocation to improve task performance while preserving well-being and safety. In this paper, we focus on the MW component of OFS and propose WAUC, an open multi-modal dataset for assessing the MW under conditions that more closely resemble real-world scenarios. More specifically, the database provides researchers with data from 48 participants, covering seven different modalities measured using off-the-shelf wearable devices, while participants performed six different MW (high/low) vs. physical workload (no/medium/high) tasks, either on a treadmill or a stationary bike. The modalities include electroencephalogram, ECG, breathing rate, skin temperature, GSR, BVP, and 3-axis accelerometry. The MATB-II assessment was used to modulate MW level. Each participant also provided subjective workload ratings using the NASA-TLX questionnaire, as well as Borg fatigue scale ratings.

Besides describing the experimental procedure, detailed validation analysis of the recorded subjective ratings and neurophysiological signals is also provided, along with a number of research directions that can be followed from the WAUC dataset. The database is available to the research community at: http://musaelab.ca/resources/.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article is available at http://musaelab.ca/resources/, without undue reservation.

ETHICS STATEMENT

Written informed consent was obtained from the individual for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

All authors: experimental design. IA, AT, MP, J-FG, DL, ST, and TF: writing and reviewing. IA, MP, AT, and RC: statistical analysis and programming. ST and TF: funding and supervision.

FUNDING

The authors wish to acknowledge funding from NSERC CRDPJ 485455-15, PROMPT Québec, MITACS, and Thales Canada.

ACKNOWLEDGMENTS

We thank Srinivasan Jayaraman for working on the experimental design, and Olivier Rosanne for feature extraction from the physiological signals.

REFERENCES

Albuquerque, I., Rosanne, O., Gagnon, J.-F., Tremblay, S., and Falk, T. H. (2019). "Fusion of spectral and spectro-temporal EEG features for mental workload assessment under different levels of physical activity," in *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)* (San Fracisco, CA), 311–314. doi: 10.1109/NER.2019.8717092

Almogbel, M. A., Dang, A. H., and Kameyama, W. (2018). "EEG-signals based cognitive workload detection of vehicle driver using deep learning," in *2018 20th International Conference on Advanced Communication Technology (ICACT)* (Chuncheon-si), 256–259. doi: 10.23919/ICACT.2018.8323716

Behar, J., Johnson, A., Clifford, G. D., and Oster, J. (2014). A comparison of single channel fetal ECG extraction methods. *Ann. Biomed. Eng.* 42, 1340–1353. doi: 10.1007/s10439-014-0993-9

Behar, J. A., Rosenberg, A. A., Weiser-Bitoun, I., Shemla, O., Alexandrovich, A., Konyukhov, E., et al. (2018). Physiozoo: a novel open access platform for heart rate variability analysis of mammalian electrocardiographic data. *Front. Physiol.* 9:1390. doi: 10.3389/fphys.2018.01390

Borg, G. A. (1982). Psychophysical bases of perceived exertion. *Med. Sci. Sports Exerc.* 14, 377–381. doi: 10.1249/00005768-198205000-00012

Borghini, G., Aricó, P., Di Flumeri, G., Sciaraffa, N., Colosimo, A., Herrero, M.-T., et al. (2017). A new perspective for the training assessment: machine learning-based neurometric for augmented user's evaluation. *Front. Neurosci.* 11:325. doi: 10.3389/fnins.2017.00325

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., and Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci. Biobehav. Rev.* 44, 58–75. doi: 10.1016/j.neubiorev.2012.10.003

Cain, B. (2007). *A Review of the Mental Workload Literature*. Technical Report, Defence Research And Development Toronto.

Camm, A. et al. (1996). Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task force of the European society of cardiology and the North American society of pacing and electrophysiology. *Circulation* 93, 1043–1065. doi: 10.1161/01.CIR.93.5.1043

Cassani, R., Banville, H., and Falk, T. H. (2015). "Mules: An open source EEG acquisition and streaming server for quick and simple prototyping and recording," in *Proceedings of the 20th International Conference on Intelligent User Interfaces Companion* (Atlanta Georgia, GA: ACM), 9–12. doi: 10.1145/2732158.2732193

Cassenti, D. N., Kelley, T. D., and Carlson, R. A. (2010). "Modeling the workload-performance relationship," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 54* (San Francisco, CA: SAGE Publications). doi: 10.1177/154193121005401968

Castaldo, R., Montesinos, L., Wan, T. S., Serban, A., Massaro, S., and Pecchia, L. (2017). "Heart rate variability analysis and performance during a repeated mental workload task," in *EMBEC & NBC 2017* (Tampere: Springer), 69–72. doi: 10.1007/978-981-10-5122-7_18

Castellanos, N. P., and Makarov, V. A. (2006). Recovering EEG brain signals: artifact suppression with wavelet enhanced independent component analysis. *J. Neurosci. Methods* 158, 300–312. doi: 10.1016/j.jneumeth.2006.05.033

Charles, R. L. and Nixon, J. (2019). Measuring mental workload using physiological measures: a systematic review. *Appl. Ergon*. 74, 221–232. doi: 10.1016/j.apergo.2018.08.028

Chaumet, G., Delaforge, A., and Delliaux, S. (2019). Mental workload alters heart rate variability lowering non-linear dynamics. *Front. Physiol*. 10:565. doi: 10.3389/fphys.2019.00565

Hart, S. G., and Staveland, L. E. (1988). "Development of NASA-TLX (task load index): results of empirical and theoretical research," in *Advances in Psychology, Vol. 52*, eds P. A. Hancock and N. Meshkati (Elsevier), 139–183. doi: 10.1016/S0166-4115(08)62386-9

Hefron, R., Borghetti, B., Schubert Kabban, C., Christensen, J., and Estepp, J. (2018). Cross-participant eeg-based assessment of cognitive workload using multi-path convolutional recurrent neural networks. *Sensors* 18:1339. doi: 10.3390/s18051339

Hockey, G. R. J. (2003a). *Operator Functional State as a Framework for the Assessment of Performance Degradation, Vol. 355*. Nato Science Series Sub Series I Life And Behavioural Sciences. IOS Press, 8–23.

Hockey, G. R. J. (2003b). *Operator Functional State: The Assessment and Prediction of Human Performance Degradation in Complex Tasks, Vol. 355*. IOS Press.

Hogervorst, M. A., Brouwer, A.-M., and Van Erp, J. B. (2014). Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. *Front. Neurosci*. 8:322. doi: 10.3389/fnins.2014.00322

Jasper, P., Sibley, C., and Coyne, J. (2016). "Using heart rate variability to assess operator mental workload in a command and control simulation of multiple unmanned aerial vehicles," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 60* (Los Angeles, CA: Sage Publications), 1125–1129. doi: 10.1177/1541931213601264

Johnson, A. M. (1990). Speed of mental rotation as a function of problem-solving strategies. *Percept. Motor Skills* 71, 803–806. doi: 10.2466/pms.1990.71.3.803

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., et al. (2011). Deap: a database for emotion analysis; using physiological signals. *IEEE Trans. Affect. Comput*. 3, 18–31. doi: 10.1109/T-AFFC.2011.15

Ladouce, S., Donaldson, D. I., Dudchenko, P. A., and Ietswaart, M. (2019). Mobile EEG identifies the re-allocation of attention during real-world activity. *Sci. Rep*. 9, 1–10. doi: 10.1038/s41598-019-51996-y

Matthews, R., Turner, P., McDonald, N., Ermolaev, K., McManus, T., Shelby, R., et al. (2008). "Real time workload classification from an ambulatory wireless EEG system using hybrid EEG electrodes," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Vancouver, BC), 5871–5875. IEEE. doi: 10.1109/IEMBS.2008.4650550

Milner, D. (1998). Cognitive neuroscience: the biology of the mind and findings and current opinion in cognitive neuroscience. *Trends Cogn. Sci*. 2:463. doi: 10.1016/S1364-6613(98)01226-1

Pan, J., and Tompkins, W. J. (1985). A real-time QRS detection algorithm. *IEEE Trans. Biomed. Eng*. 32, 230–236. doi: 10.1109/TBME.1985.325532

Reid, G. B., and Nygren, T. E. (1988). "The subjective workload assessment technique: a scaling procedure for measuring mental workload," in *Advances in Psychology, Vol. 52*, eds P. A. Hancock and N. Meshkati (Elsevier), 185–218. doi: 10.1016/S0166-4115(08)62387-0

Ruffini, G., Dunne, S., Farrés, E., Cester, Í., Watts, P. C., Ravi, S., et al. (2007). "ENOBIO dry electrophysiology electrode; first human trial plus wireless electrode system," in *Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Lyon), 6689–6693. doi: 10.1109/IEMBS.2007.4353895

Santiago-Espada, Y., Myer, R. R., Latorella, K. A., and Comstock J. R. Jr. (2011). *The Multi-Attribute Task Battery II (MATB-II) Software for Human Performance and Workload Research: A User's Guide*. Available online at: NASA's website https://ntrs.nasa.gov/citations/20110014456

Sauer, J., Wastell, D., and Hockey, G. (2000). A conceptual framework for designing micro-worlds for complex work domains: a case study of the cabin air management system. *Comput. Hum. Behav*. 16, 45–58. doi: 10.1016/S0747-5632(99)00051-5

Shepard, R. N., and Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science* 171, 701–703. doi: 10.1126/science.171.3972.701

Shu, L., Xu, T., and Xu, X. (2019). Multilayer sweat-absorbable textile electrode for EEG measurement in forehead site. *IEEE Sens. J*. 19, 5995–6005. doi: 10.1109/JSEN.2019.2912667

Snyder, K. L., Kline, J. E., Huang, H. J., and Ferris, D. P. (2015). Independent component analysis of gait-related movement artifact recorded using EEG electrodes during treadmill walking. *Front. Hum. Neurosci*. 9:639. doi: 10.3389/fnhum.2015.00639

Soleymani, M., Lichtenauer, J., Pun, T., and Pantic, M. (2011). A multimodal database for affect recognition and implicit tagging. *IEEE Trans. Affect. Comput*. 3, 42–55. doi: 10.1109/T-AFFC.2011.25

Teplan, M. et al. (2002). Fundamentals of EEG measurement. *Measure. Sci. Rev*. 2, 1–11. Available online at: https://www.measurement.sk/2002/S2/p2.html

Ting, C.-H., Mahfouf, M., Nassef, A., Linkens, D. A., Panoutsos, G., Nickel, P., et al. (2009). Real-time adaptive automation system based on identification of operator functional state in simulated process control operations. *IEEE Trans. Syst. Man Cybern. Part A* 40, 251–262. doi: 10.1109/TSMCA.2009.2035301

Vidullch, M. A., Ward, G. F., and Schueren, J. (1991). Using the subjective workload dominance (sword) technique for projective workload assessment. *Hum. Fact*. 33, 677–691. doi: 10.1177/001872089103300605

Wierwille, W. W., and Casali, J. G. (1983). "A validated rating scale for global mental workload measurement applications," in *Proceedings of the Human Factors society Annual Meeting, Vol. 27* (Norfolk, VA: Sage Publications), 129–33. doi: 10.1177/154193128302700203

Wilson, G. F., and Russell, C. A. (2003a). Operator functional state classification using multiple psychophysiological features in an air traffic control task. *Hum. Fact*. 45, 381–389. doi: 10.1518/hfes.45.3.381.27252

Wilson, G. F., and Russell, C. A. (2003b). Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. *Hum. Fact*. 45, 635–644. doi: 10.1518/hfes.45.4.635.27088

Yin, Z., and Zhang, J. (2017). Cross-session classification of mental workload levels using EEG and an adaptive deep learning model. *Biomed. Signal Process. Control* 33, 30–47. doi: 10.1016/j.bspc.2016.11.013

Zhang, J., Yin, Z., and Wang, R. (2016). Pattern classification of instantaneous cognitive task-load through GMM clustering, Laplacian Eigenmap, and ensemble SVMs. *IEEE/ACM Trans. Comput. Biol. Bioinform*. 14, 947–965. doi: 10.1109/TCBB.2016.2561927

Zink, R., Hunyadi, B., Van Huffel, S., and De Vos, M. (2016). Mobile EEG on the bike: disentangling attentional and physical contributions to auditory attention tasks. *J. Neural Eng*. 13:046017. doi: 10.1088/1741-2560/13/4/046017

Conflict of Interest: J-FG and DL were employed by the company Thales Research and Technology Canada.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX

TABLE A1 | Subjective ratings descriptive statistics (mean and standard deviation) for subjects that used the treadmill (top rows) and bike (bottom rows) during the experiment.

		No PW		Medium PW		High PW	
		Low MW	**High MW**	**Low MW**	**High MW**	**Low MW**	**High MW**
Treadmill							
NASA-TLX	Mental demand	8.41 ± 5.66	11.27 ± 6.63	10.36 ± 5.09	13.41 ± 5.92	11.23 ± 5.07	15.50 ± 3.56
	Physical demand	4.32 ± 5.05	5.36 ± 5.66	8.23 ± 4.89	9.00 ± 5.12	14.59 ± 4.75	15.41 ± 4.54
	Temporal demand	7.23 ± 6.05	10.27 ± 6.91	8.41 ± 4.59	11.68 ± 5.91	11.41 ± 5.48	15.05 ± 4.34
	Performance	16.36 ± 4.10	12.27 ± 4.12	14.95 ± 4.13	11.05 ± 3.90	15.14 ± 4.11	12.14 ± 4.28
	Effort	8.86 ± 5.76	11.68 ± 5.56	11.45 ± 4.64	13.64 ± 5.19	13.91 ± 4.51	16.36 ± 3.55
	Frustration	6.64 ± 6.64	8.64 ± 6.77	6.59 ± 5.75	10.32 ± 7.17	7.86 ± 6.56	10.18 ± 6.73
Borg Scale	Before break	8.05 ± 2.77	13.59 ± 2.77	10.09 ± 3.04	11.36 ± 2.98	8.86 ± 3.21	14.77 ± 2.16
	After break	8.95 ± 3.20	12.64 ± 2.85	9.27 ± 2.37	10.00 ± 2.62	8.64 ± 3.33	12.95 ± 3.11
Bike							
NASA-TLX	Mental demand	6.40 ± 3.44	10.12 ± 4.00	9.00 ± 4.71	11.36 ± 4.70	9.24 ± 4.99	12.60 ± 4.56
	Physical demand	3.04 ± 2.59	3.44 ± 3.22	7.96 ± 3.79	8.84 ± 4.79	11.00 ± 5.40	12.92 ± 4.81
	Temporal demand	5.44 ± 3.48	9.20 ± 4.02	8.20 ± 4.53	11.20 ± 5.37	9.56 ± 4.84	12.80 ± 4.86
	Performance	17.20 ± 4.01	12.60 ± 4.43	15.20 ± 4.07	12.72 ± 4.43	15.08 ± 4.56	12.88 ± 4.56
	Effort	6.48 ± 4.11	11.20 ± 4.05	10.48 ± 4.48	12.52 ± 5.12	11.12 ± 5.37	13.40 ± 4.44
	Frustration	4.28 ± 3.25	6.88 ± 5.09	6.00 ± 4.12	9.12 ± 6.02	8.60 ± 5.39	8.56 ± 4.71
Borg Scale	Before break	7.20 ± 1.58	12.40 ± 2.69	10.36 ± 2.02	11.24 ± 2.73	8.56 ± 2.53	12.92 ± 2.66
	After break	6.96 ± 1.24	11.16 ± 2.58	8.92 ± 2.10	10.20 ± 2.71	8.00 ± 2.18	11.32 ± 2.58

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 08 December 2020
doi: 10.3389/fnins.2020.542934

PASS: A Multimodal Database of Physical Activity and Stress for Mobile Passive Body/ Brain-Computer Interface Research

Mark Parent[1], *Isabela Albuquerque*[1], *Abhishek Tiwari*[1], *Raymundo Cassani*[1], *Jean-François Gagnon*[2], *Daniel Lafond*[2], *Sébastien Tremblay*[3] *and Tiago H. Falk*[1,4]*

[1] INRS-EMT, Université du Québec, Montréal, QC, Canada, [2] Thales Research and Technology Canada, Quebec City, QC, Canada, [3] Université Laval, School of Psychology, Quebec City, QC, Canada, [4] PERFORM Center, Concordia University, Montréal, QC, Canada

OPEN ACCESS

Edited by:
Davide Valeriani, Massachusetts Eye and Ear Infirmary and Harvard Medical School, United States

Reviewed by:
Stephen Fairclough, Liverpool John Moores University, United Kingdom
Antonio Fernández-Caballero, University of Castilla-La Mancha, Spain
Sujoy Ghosh Hajra, National Research Council Canada (NRC-CNRC), Canada

***Correspondence:**
Tiago H. Falk
falk@emt.inrs.ca

Specialty section:
This article was submitted to Neural Technology, a section of the journal Frontiers in Neuroscience

Received: 14 March 2020
Accepted: 16 November 2020
Published: 08 December 2020

Citation:
Parent M, Albuquerque I, Tiwari A, Cassani R, Gagnon J-F, Lafond D, Tremblay S and Falk TH (2020) PASS: A Multimodal Database of Physical Activity and Stress for Mobile Passive Body/ Brain-Computer Interface Research. Front. Neurosci. 14:542934. doi: 10.3389/fnins.2020.542934

With the burgeoning of wearable devices and passive body/brain-computer interfaces (B/BCIs), automated stress monitoring in everyday settings has gained significant attention recently, with applications ranging from serious games to clinical monitoring. With mobile users, however, challenges arise due to other overlapping (and potentially confounding) physiological responses (e.g., due to physical activity) that may mask the effects of stress, as well as movement artifacts that can be introduced in the measured signals. For example, the classical increase in heart rate can no longer be attributed solely to stress and could be caused by the activity itself. This makes the development of mobile passive B/BCIs challenging. In this paper, we introduce PASS, a multimodal database of Physical Activity and StresS collected from 48 participants. Participants performed tasks of varying stress levels at three different activity levels and provided quantitative ratings of their perceived stress and fatigue levels. To manipulate stress, two video games (i.e., a calm exploration game and a survival game) were used. Peripheral physical activity (electrocardiography, electrodermal activity, breathing, skin temperature) as well as cerebral activity (electroencephalography) were measured throughout the experiment. A complete description of the experimental protocol is provided and preliminary analyses are performed to investigate the physiological reactions to stress in the presence of physical activity. The PASS database, including raw data and subjective ratings has been made available to the research community at http://musaelab.ca/pass-database/. It is hoped that this database will help advance mobile passive B/BCIs for use in everyday settings.

Keywords: neurophysiology, wearables, physical activity, stress, body/brain-computer interfaces, heart rate variability, electroencephalography (EEG), electrodermal activity

1. INTRODUCTION

Brain-computer interfaces (BCIs) are systems that provide communication and control abilities to users without relying on the brain's normal output pathways (Wolpaw et al., 2000). BCIs are typically divided into two categories (Tan and Nijholt, 2010): active or passive. Active BCIs are systems where users must actively modulate their brain responses in order to control the BCI. Passive BCIs, in turn, monitor the user's implicit states, thus do not require the user to perform any specific task. More recently, some researchers have started to use the term "Body/brain-computer interfaces" (B/BCIs) to extend the inputs of BCI to the rest of the physiological system (e.g., Feng et al., 2016).

Physiological measures and passive body/brain-computer interfaces offer tremendous possibilities for monitoring individual functional states. In recent years, several works have shown that physiological measures can be used to assess e.g., the operator functional state of workers (i.e., workload, stress, fatigue), videogame player fun level, or even health markers (Banaee et al., 2013; Gagnon et al., 2016; Harrivel et al., 2017; Fortin-Côté et al., 2018). Moreover, it has been demonstrated that such assessment can be leveraged to augment interactions with intelligent systems, such as adaptive videogames or adaptive workload management systems (Parnandi and Gutierrez-Osuna, 2015; Aricò et al., 2016). Wearables further push this progress by increasing portability and accessibility of neurophysiological measures, while reducing the cost associated with such systems.

There are many challenges, however, with relying on neurophysiological measures and passive B/BCIs in realistic settings where the user is mobile and multi-tasking. The first relates to the question of multidimensionality of psychological states (Matthews et al., 2015) where different emotions and psychological conditions are combined. While the multidimensionality of psychological states can be well-captured with questionnaires, it becomes harder with metrics derived from neurophysiological models. One example of this is the overlapping of e.g., physical activity and stress on heart rate and heart rate variability. An additional challenge lies on the artifacts that are generated once experiments are performed outside controlled laboratory settings with sensors that are sensitive to e.g., movement artifacts (Sun et al., 2010; Falk et al., 2016).

The first goal of this project is to provide a multimodal dataset where affective stress and physical activity are both modulated. To date, there are no publicly-available datasets that explore the concurrent modulation of affective stress and physical activity and the impact it has on physiological measures and on artifact generation. We aim to fill this gap. The second goal of this article is to provide a dataset that mimics realistic settings to support "in-the-wild" B/BCI development. To do so, we used a realistic task setting (i.e., playing video games) and used off-the-shelf wearable devices. Modalities used in this study include electroencephalography, cardiac activity, electrodermal activity, breathing information, and skin temperature.

In this paper, we describe PASS, a multimodal database of Physical Activity and StresS. Here, we present the experimental protocol used, descriptive statistics of the recorded neurophysiological signals under the varying conditions, and also introduce preliminary results on the use of machine learning to model stress that is robust to different physical activity confounding factors. The database has been made publicly available at http://musaelab.ca/pass-database/, along with stress and physical fatigue questionnaire responses provided by the participants.

In the remainder of this paper, we first provide background on the theory and physiological measures of stress in section 2, followed by a description of the current challenges in stress monitoring in section 3. Next, a full description of the experimental design and the methodology used to perform the data collection is presented in section 4. Validation of the dataset is presented in section 5, including analyses on the physiological and subjective data gathered. Results are then discussed in section 6 and conclusions drawn in section 7.

2. BACKGROUND

2.1. Theory of Stress

Stress is a psychological concept that has received a tremendous level of scientific attention throughout its history. One could argue that this attention is well-placed, as stress is well-known to have several negative effects on individual health and performance. While many definitions of stress exist, it can be generally defined as an ensemble of coping responses to react to a perceived threat (Lazarus and Folkman, 1984).

While some amount of stress is inevitable, extended or acute exposure to stress is known to be associated with several health problems such as cardiovascular diseases, respiratory diseases, and autoimmune diseases (Schneiderman et al., 2005). Investigations of occupational stress in many countries have shown that a large proportion of the population is exposed to detrimental levels of stress through their work environment (Jones et al., 2016), increasing absenteeism and turnover intention (Jamal, 2007). Finally, stress is associated with psychological disorders like depression (Caspi et al., 2003).

Besides health considerations, several researchers have described intricate links between stress and human performance. Stress has been shown to influence cognitive performance, such as memory. Authors suggest that high arousal could enhance memory consolidation, but could hamper memory recall (Wolf, 2009). Anxiety is also linked with poorer manual dexterity (Kneller et al., 2012; Skirbekk et al., 2012). In job settings, stress is associated with lower job performance (Jamal, 2007). Despite these results, some findings suggest that stress might be beneficial in some circumstances. Using a crisis management simulation, authors investigated the link between stress (i.e., time pressure) and team communication. They found that stress increases communication quantity and efficiency. They do, however, underline that frequent requests for information are associated with poorer task performance (Pfaff, 2012). Stress also influences academic performance. A recently published longitudinal study showed that children and adolescents undergoing an anxiety treatment therapy were associated with better academic performance (Swan et al., 2018).

In the literature, stress is conceptualized in various ways. First, stress can refer to shorter-term activation, caused by more immediate situations (e.g., solving a problem). On the other hand, stress can also relate to longer-term straining states (i.e., chronic stress, occupational stress), caused by adverse life or job situations (e.g., disease, mourning, layoff) or by prolonged exposure to short-term stress (Schneiderman et al., 2005; Schubert et al., 2009). In experimental settings, most researchers use *validated stressors* to induce stress in participants. For example, the cold-pressor test which requires participants to submerge limbs in near-frozen water for a short period of time has been used in several studies investigating stress (McRae et al., 2006; Duncko et al., 2009; Dierolf et al., 2017).

Validated stressors can elicit two forms of stress: mental stress and affective stress. Mental stress refers to situations that require reflection and problem-solving abilities (Sun et al., 2010; Al-Shargie et al., 2016). For example, the stroop task or mental arithmetic task, are designed to stress individuals by requiring mental effort (Visnovcova et al., 2014). Mental stress is closely tied to the concept of mental workload. Mental workload can be difficult to define (Young et al., 2015). In general, it can be considered as the level of mental resources required to meet a specific performance (Young et al., 2015). On the other hand, affective stress relates to anxiety, fear or discomfort. Such stressors include the Trier Social Stress Task (Kudielka et al., 2007). The Trier Social Stress Task requires individuals to perform tasks, such as oral presentations or mental arithmetic, in front of fake experts. Affective stress is generally associated with emotions of negative valence (Hwang et al., 2018), with various levels of arousal. Therefore, affective stressors also include viewing emotionally loaded stimuli, such as pictures, movies, or sentences (Wolf, 2009).

2.2. Physiological Measures of Stress

Stress can be assessed using subjective measures. Various questionnaires have been developed to measure stress related to tasks (Matthews and Campbell, 2010) or anxiety (Spielberger, 2010). Subjective measures have the advantage of being simple and to offer direct access to cognition; however, they are also known to be biased. Furthermore, they require interruptions. Physiological measures, on the other hand, are objective and can be taken continuously, without interruptions. As such, several recent studies have proposed physiology-based models, sometimes achieving fairly high detection accuracy (Smets et al., 2019).

2.2.1. Neurophysiological Measures

Stress generates a wide range of physiological reactions that can be leveraged to measure its intensity in individuals. It can be assessed using electroencephalography (EEG), but elicited patterns are very dependant on the type of stressor used. Task demand and temporal pressure are often associated with a decrease in the alpha band power in various cerebral regions, including frontal, central and parietal, and associated with an increase in theta at frontal and parietal regions (Borghini et al., 2014; Al-Shargie et al., 2016). Individuals performing the Montreal Imaging Stress Task (mental arithmetic combined with negative social feedback) have been shown to exhibit greater relative gamma band power in prefrontal, temporal, and parietal regions (Minguillon et al., 2016). Similarly, the gamma band is associated with worry. Individuals suffering from generalized anxiety disorder undergoing a worry task (self-selected worrying thought) exhibited greater gamma power in temporal and parietal lobes (Oathes et al., 2008). In another study, prefrontal asymmetry of participants performing a virtual reality surveillance task was investigated (Brouwer et al., 2011). During stressful moments (i.e., a bomb explosion combined with negative feedback), alpha asymmetry of prefrontal regions (F7-F8) was significantly higher than during non-stressful moments. Prefrontal asymmetry was also associated with stress in other studies, such as participant performing the Maastricht Acute Stress Task (Quaedflieg et al., 2015). While not investigated directly in studies involving stress, amplitude modulation features of EEG have shown discriminative power for valence and arousal measurement (Clerico et al., 2018), as well as workload (Albuquerque et al., 2018). Stress is also known to influence event-related potentials, for example, during sustained attention tasks (Righi et al., 2009). Apart from EEG, stress can be measured using other neurophysiological measures, such as functional near-infrared spectroscopy (Al-Shargie et al., 2016; Parent et al., 2019a).

2.2.2. Cardiac Measures

Stress is well-known to increase heart rate. Heart rate is often derived from the electrocardiography (ECG) signal. ECG consists of placing electrodes on the skin to measure the voltage difference caused by the electrical activity of the heart. Heart rate can also be measured using photoplethysmography by measuring variations of the light absorption of the skin. Apart from heart rate, stress is known to influence heat-rate variability (Kreibig et al., 2007; Castaldo et al., 2015). Heart rate variability is the analysis of the changes in heart rhythm. Heart rate variability does not usually refer to a specific feature, but a family of features, each describing various aspects of cardiac activity. As such, stress is known to increase the standard deviation of inter-beat intervals (SDNN) or reduce the root mean square of inter-beat intervals (RMSSD) (Castaldo et al., 2015). Stress also influences frequency-domain features of heart rate variability, as the ratio between low and high frequency power (Castaldo et al., 2015). Blood pressure is also influenced by stress. Fear is known to increase both systolic and diastolic blood pressure (Kreibig et al., 2007). In a simulation of computer work containing stressful and non-stressful sessions, it was shown that blood pressure increased during work sessions compared to rest, but did not decrease during non-stressful sessions (Hjortskov et al., 2004).

2.2.3. Breathing Measures

Breathing rate increases under stress (Rainville et al., 2006; Homma and Masaoka, 2008). Furthermore, anxious individuals tend to breathe faster during anticipatory stress than less anxious individuals (Homma and Masaoka, 2008). Studies have also shown that respiratory variability is higher and more random during mental stress and worry (Vlemincx et al., 2013). In the same line of thought, fear is associated with higher standard

deviation of breathing amplitude (Rainville et al., 2006). Sighing seems more present during stress (Vlemincx et al., 2013). It is suggested that sighing might act as a reset to irregular respiration pattern encountered during stress.

2.2.4. Electrodermal Measures

Stress also has effects on sweating, which can be measured using electrodermal activity (EDA). EDA is described as the electrical conductance of the skin, which is modulated by the level of sweat. Sweat is well-known to be influenced by physical activity. However, it is suggested that sweat glands are controlled by the sympathetic system. EDA is thus considered as a proxy to observe the sympathetic activation of individuals. Besides the electrodermal level (i.e., the "amount" of sweat on the skin), EDA can be described in greater details by analyzing electrodermal responses. Electrodermal responses are brief "peaks" of sweat that occurs in response to a stimulus. They can be specific (i.e., related to a known event) or non-specific (Boucsein, 2012). Typically, short-term stressors used in laboratory settings, such as the cold pressor or stroop task, or fearful states tend to increase electrodermal level, non-specific response frequency, as well as response amplitude (Kreibig et al., 2007; Reinhardt et al., 2012; Posada-Quintero et al., 2016b, 2018a). While still not very common, some authors have investigated frequency domain features of the EDA. Overall, results suggest that the stressors influence mostly the 0.045 to 0.15 Hz band (Posada-Quintero et al., 2016a). Frequency domain features of EDA are said to be sometimes more sensitive to stress than classical time-domain features (Posada-Quintero et al., 2016b, 2018a).

2.2.5. Thermal Measures

In reaction to stress, mammals, including humans, typically have a reduced temperature in peripheral regions, while the temperature of the face and core region rises (Marazziti et al., 1992; Vianna and Carrive, 2005; Kreibig et al., 2007; Nakamura, 2011). It is theorized that this reaction is caused by a constriction of the peripheral arterioles, which could reduce blood loss if a wound occurred.

3. CURRENT CHALLENGES

3.1. Multidimensionality of Stress

It is challenging to fully separate mental stress from affective stress, as all mental tasks will still trigger even a low amount of anxiety in individuals. Conversely, affective stress will probably trigger even a small amount of mental activity, whether it is due to assessing the threat, planning a response or simply diverting attention to less stressful states. Yet, both types of stress have different implications. For example, authors suggest that mental forms of stress (like engagement) correlate with working memory performance while affective forms of stress (like distress) negatively correlate with performance (Qin et al., 2009; Matthews and Campbell, 2010). On the physiological level, it is suggested that mental effort is associated with sympatho-adrenal-medullary axis (epinephrine and norepinephrine) while affective stressors are more associated with the hypothalamus-pituitary-adrenal (cortisol) axis. While there is not an extensive amount of literature to support this, it can be surmised that high mental stress with minimal affective stress might lead to positive outcomes (like task completion) while high affective stress without much mental activation is not beneficial in any way. This view was supported by some authors investigating physiological differences between mental effort and distress (Frankenhaeuser, 1986; Gaillard and Wientjes, 1994; Matthews et al., 2015) and does, to a certain extent, resemble the eustress/distress dissociation proposed by Hans Seyle in his classical work on stress (Selye, 1985).

Subjective tools attempt to distinguish between these nuances of stress. The NASA-TLX questionnaire, for example, features a "Frustration" axis, covering affective load among more cognitive ones (Hart, 2006). The Dundee Stress State Questionnaire also distinguishes more mental stress (i.e., engagement) from affective forms (i.e., distress, worry) (Matthews and Campbell, 2010). In contrast, physiological measures of stress, despite being well-documented, are rarely interpreted in a multidimensional way (Matthews et al., 2015). Distinguishing mental and affective stress using physiology remains a challenge today. The separation of mental and affective stress goes beyond the scope of this database description work, thus henceforth, the term "stress" will be used to comprise their combined effects. Notwithstanding, future work can explore such separation with multimodal tools (e.g., Parent et al., 2019b).

3.2. Stress Detection in Laboratory and Ambulatory Settings

Given the numerous effects of stress on the human body, research has focused on trying to propose models to detect stress based on physiology. In a recent article (Smets et al., 2019), the authors reviewed 25 papers that investigated this research question over the last several years. Comparing the performance of each model investigated in these studies can be difficult as several factors can differ between studies. First, as detailed previously, different stressors can be used. Second, models use different physiological modalities and, in some cases, different combinations of modalities. Models also differ in terms of classification scheme (i.e., within participants, between participants) and classification levels. Most studies propose models that distinguish between a resting state and a stressful task. However, some studies attempt to classify multiple levels of stress (e.g., low, medium, high) and others use regression models to measure a stress level (e.g., Hovsepian et al., 2015). Finally, as described by Smets et al. (2019), the majority of papers focus on laboratory settings, while only a select few have attempted to detect stress in ambulatory settings.

In laboratory settings, classification accuracy of stress detection models can reach fairly high levels. In a recent example, researchers used a portable wristband, recording heart and electrodermal activity, to detect affective stress induced by the Trier Social Stress Task. They reported achieving an area under the receiver operating characteristic curve of 0.87 (Ollander et al., 2016). In another case, researchers used ECG and EEG to classify the affective state of individuals playing a survival horror game (Vachiratamporn et al., 2013). Six different affective states were

classified. Authors reported up to 90% classification accuracy using ECG and up to 73% using EEG. In a recent study, the Muse headband (i.e., the same low-cost EEG system used in this study) was used to classify the subjective stress level of participants (Arsalan et al., 2019). Authors reported accuracy as high as 92% on a two-class classification task. Finally, by using EEG and near-infrared spectroscopy, detection of mental stressors with accuracy near 95% (i.e., distinguishing between control and stress) has been reported in Al-Shargie et al. (2016).

In ambulatory settings, however, performance is usually lower. Nonetheless, the topic gained scientific attention in the last few years, improving the potential of ambulant stress detection models. In two recent examples, portables sensors (i.e., a chest strap, a wristband) were used to detect stress in ambulatory settings (Hovsepian et al., 2015; Gjoreski et al., 2016). Models reached, respectively, a 0.72 correlation coefficient or 76% classification accuracy at detecting self-reported stress (i.e., two-class classification). In a recent study, EEG asymmetry was used to monitor arousal and valence of individuals in the presence of physical activity (i.e., construction workers) (Hwang et al., 2018). While comparison to ground truth is difficult in naturalistic situations, the authors suggest that this method had potential to assess emotional state of individuals, especially for valence detection.

Despite advances in ambulant stress detection, several challenges are still in the way of a highly robust stress detection model. As suggested by Smets et al. (2019), movement and physical activity are the most obvious limitations of stress detection models. Some models are configured to not predict stress if physical activity is detected (Hovsepian et al., 2015), thus not only biasing error rate measures, but also preventing stress detection in the presence of physical activity. Other models are configured to receive contextual data (such as physical activity), improving accuracy in exchange of manual input in the model (Gjoreski et al., 2016). However, very few papers have investigated stress detection in the presence of varying levels of physical activity.

Movement and physical activity affect physiological measures in three different ways. First, physiological measures are influenced by the direct consequences of physical activity. When individuals start to perform physical activity, the body triggers a series of physiological mechanisms to shift from a rest state to an active state. The most obvious example is the increase in heart rate caused by physical exertion (Bernardi et al., 1996). Since physical activity requires energy, the heart must beat faster to deliver more supplies to muscle cells, fetch more oxygen and reject more CO_2 in the lungs. The skin sweat will also be increased to dissipate excess of heat caused by physical activity (Neto et al., 2010). The response of the central system will also be affected, as some areas of the brain will be required to coordinate limb movements.

Second, physiological measures are influenced by shifts in psychological states that come with physical activity. For example, it has been shown that performing physical tasks, such as lifting boxes, will draw mental resources (DiDomenico and Nussbaum, 2008). In relation with this paper, there is also some evidence that physical activity can reduce long term stress and anxiety (Pedersen and Saltin, 2015). While scientific attention is mostly oriented toward long-term benefits of physical exercise regarding stress, evidence also suggests the presence of short-term effects (Salmon, 2001). Individuals are most likely to report having a better mood immediately after exercise. Some factors modulate this relationship. Having a poor mood before exercise usually causes a sharper improvement in mood after exercise. On the other hand, performing at higher intensity than habitual level can deteriorate mood.

Finally, movement and physical activity alter physiological recordings through noise or signal loss. If the device uses electrodes (e.g., EEG, ECG, EDA), these might lose contact with the skin, briefly or continuously, and alter the measured signal (Castellanos and Makarov, 2006; Gwin et al., 2010). The nature of the physical task might also displace, disable or even damage sensors. If the data is transmitted wirelessly, signal loss might be encountered when the distance between the emitter and the receiver is too high or when an obstacle is present between them.

4. METHODS AND MATERIALS

4.1. Motivation and Overview

The experiment discussed in this paper sought to elicit affective stress. Common stressors used in psychophysiology (e.g., Stroop task, n-back task) were excluded since they were not sufficiently independent from mental stress. In a similar way, time pressure (sometime used as a stressor) was also discarded since higher time pressure can sometimes lead to higher mental effort. To support "in-the-wild" B/BCI development, we also sought to use a realistic task setting. Therefore, a survival video game was selected as a stressor. Video games have already been used in affect research, and, as in our case, in combination with physiological measures. Survival video games also allow for a short-duration experimental design (compared to studies that focus on more chronic, long-term stressors).

More specifically, the experiment consisted of playing video games while pedaling on a stationary bike. Two experimental variables were manipulated: stress and physical activity intensity. There were two stress levels (no stress/stressful) and three physical activity levels (0, 18, 24 km/h). Participants performed all six combinations in counterbalanced order. Each trial lasted 10 min. Physiological activity and subjective ratings were recorded throughout the experiment. The following sections will provide more details about the experimental design.

4.2. Stress Manipulation

Stress was modulated by switching between two video games: a non-stressful one, serving as a control condition, and a stressful one. The non-stressful game used was TIMEframe. TIMEframe is a commercially available exploration/puzzle game developed by Random Seed Games (Random Seed Games, 2015). In TIMEframe, players must explore ruins of an abandoned city and find artifacts. The game is played from a first-person perspective and controls are similar to other first-person games. Several elements made TIMEframe a prime choice for a non-stressful game. First, there are no significant threats in the game, as the players' personas can not be harmed or die. Also, the music

FIGURE 1 | Screenshot from TIMEframe.

is soft and the environment is bright and peaceful. To further decrease stress, players were told that the number of artifacts they found would not matter and would not be recorded. The game was controlled with an Xbox One controller. **Figure 1** shows a screenshot of the game.

On the other hand, the stressful game used was Outlast. Outlast is a commercially available survival game developed by Red Barrels (Red Barrels Games, 2013). Like TIMEframe, Outlast is viewed from a first-person perspective and controlled in a similar fashion (albeit, slightly more complex than TIMEframe). The goal of the game is to navigate in a creepy asylum and evade capture/harm by its dangerous inmates. In Outlast, players cannot fight, they can only avoid, escape or hide from enemies. The game features several elements to increase stress, such as an eerie music/sound design and a horror-style environment. Some in-game areas are also poorly lit, requiring players to use a limited night vision mode. The experiment room ambient light was also dimmed to further increase stress. Outlast is deterministic and features a fairly linear playthrough, increasing the similitude of experience between participants. Once again, the game was played with an Xbox One controller. **Figure 2** depicts two screenshots of the game, one showing normal and the other (bottom) night vision mode.

Video games have been used in psychophysiological studies for some years (e.g., Carroll et al., 1987). Still, their potential as stressors is fairly unexplored. One study found that video games can elicit similar effects to common stressors such as the Trier Social Stress Task (Guitard et al., 2010). On the other hand, some studies have suggested that they have the potential to stress individuals in a way that differ from commonly used stressors, calling for more investigation on the subject (Porter and Goolkasian, 2019). The TIMEframe/Outlast manipulation was designed with stress in mind. However, this manipulation might have elicited other aspects of cognition, such as workload, engagement or enjoyment.

4.3. Physical Activity Manipulation

Physical activity was induced by asking participants to pedal on a stationary bike. The bike used featured an adjustable seat, a resistance setting and a display. The resistance was set to its minimum value (no resistance) to maximize reproducibility of the experiment. Since the participant held a controller throughout the experiment, the bike handles were unused and were flipped (see **Figure 3**). The bike display was set to show speed (in km/h).

Physical activity was modulated by changing the required speed at which participants pedaled on the bike. Our ultimate goal was to induce physical activity and artifacts to the sensors, though not to a point of making the data completely unusable. Therefore, three levels of physical activity were used. In the first level, the speed was 0 km/h; participants were simply told to sit on the bike and not pedal. At the second level, participants were told to maintain a target speed of 18 km/h. At the third level, the target speed was raised to 24 km/h. During a pilot study, we found that these speeds provided the optimal trade-off in signal quality and movement artifact generation. While most of the physical effort was made by the legs of participants, the fact that they had to hold the controller in their hand inevitably created head sways and movements. Since it can be difficult to maintain a constant speed, a tolerance of ± 2 km/h was allowed.

FIGURE 2 | Screenshot from Outlast showing normal **(top)** and night vision modes **(bottom)**.

Experimenters warned participants who drifted from the target speed during the trials. Despite focusing on the video games, the pilot study showed that most participants were able to maintain speed within the tolerance levels.

4.4. Counterbalancing

Each participant completed all six combinations of stress (no stress, stressful) and physical activity (0, 18, 24 km/h). The order of these conditions was counterbalanced and pseudo-randomized. All conditions from the same video game were performed subsequently. This was designed to avoid constant psychophysiological shifts between calm and stressful states. Doing so also allowed participants to learn the controls of one video game at a time instead of two. Each condition lasted 10 min. In TIMEframe, there were no differences in the three times participants played the game except that participants were told not to seek the same artifacts as previous sessions. For Outlast, a different scene (start point) was selected for each of the three times that participants played the game. **Table 1** describes the three in-game start points. While it is technically possible

FIGURE 3 | Experimental setup from the front **(left)**. Experimental setup from the back **(right)**. BioHarness 3 not shown since worn under the shirt.

TABLE 1 | Description of the three starting points used in Outlast.

Scene	In-game description
Admin	After being thrown out the window (admin block).
Ward	After waking up in the male ward cell.
Sewers	After the valve puzzle (chased by Chris).

for a participant to reach another condition start point before finishing the conditions, they were sufficiently distanced, so it never happened for any participant.

4.5. Physiological Measures

As mentioned earlier, one of the goals of this study is to provide a database that is captured using off-the-shelf devices. Four wearable physiological devices were used in this study. A BioHarness 3 was used to measure cardiac and respiratory activity. The BioHarness 3 is a chest strap worn directly on the skin. It measures heart activity through ECG at a sampling frequency of 250 Hz. Respiration is recorded by measuring the extension of the chest strap (18 Hz). Besides ECG and respiration, 3-axis acceleration (100 Hz) is also recorded by the device (these signals were not used in the current study). An E4 wristband was also used. The E4 records blood volume pulse through photoplethysmography (64 Hz), as well as skin temperature (4 Hz). Two electrodes, located inside the bracelet, also record galvanic skin responses (4 Hz). Cerebral activity was recorded using a Muse headband. This headband records EEG activity using 4 electrodes (TP9, AF7, AF8, and TP10) with reference to Fpz, at a 220 Hz sampling rate. From our past experience with the Muse headband, we have found that re-referencing the signals to electrodes over the temporal lobes (TP9 or TP10) could negatively impact the EEG recordings, as these signals are more prone to movement artifacts. Therefore, the acquired EEG signals were not re-referenced prior to analyses. The BioHarness 3, E4 and Muse data were streamed to a nearby laptop using Bluetooth protocol. Data was recorded using the MuSAE Lab EEG Server (MuLES), which was also used to send triggers marking the beginning and end of trials (Cassani et al., 2015).

4.6. Subjective Measures

Beside physiological measures, subjective measures were also collected. Two questionnaires were used: the NASA-TLX and the BORG. NASA-TLX is a questionnaire designed to measure workload of individuals. The original version features six questions, which must be answered on a 21-point Likert scale. In this experiment, two additional questions related to stress and

TABLE 2 | Stress and fear questions added to the NASA-TLX questionnaire.

Label (English)	Question (English)
Stress	How stressful was the task?
Fear	How scary was the task?
Label (French)	**Question (French)**
Stress	À quel point la tâche était-elle stressante?
Peur	À quel point la tâche était-elle effrayante?

fear were added to suit the research questions of the project. **Table 2** shows the extra questions used. These questions were asked in French to all participants who spoke French as their first language.

4.7. Experimental Procedure

Forty-eight participants were invited to perform an experiment at Université Laval (Quebec City, Canada). Participants were recruited using mailing lists. Candidates with heart or respiratory problems or having neurological/psychological disorders were excluded from the experiment. Given the nature of the stressor, precautions were taken to make sure participants were comfortable playing Outlast. People with a history of aversive reaction to horror (e.g., panic attacks, related phobia or just unease with featured themes) were excluded from the experiment. To avoid bias, participants who played either TIMEframe or Outlast in the past could not participate in the study. During the tutorial, participants were given warning about the expected features of the stressor. The tutorial reminded participants that they could interrupt their involvement at any moment without prejudice. Experimenters were also trained to check participant's well-being during Outlast's practice, game sessions and breaks. The experimental protocol was approved by the Ethics Review Boards of the Institut national de la recherche scientifique (INRS; Reference number: CER-16-425), the PERFORM Center (Concordia University; Reference number: 30006772) and Université Laval (Reference number: 2016-274). Participants gave written consent to participate in the study and were remunerated for their time.

Participants were greeted and invited to fill a consent form and demographic questionnaires. After these, they were briefed on the experimental procedure. Once done, physiological sensors were donned and configured in a particular order. The BioHarness 3 chest-strap (Zephyr, USA) and the E4 wristband (Empatica, USA) were donned first on the participant, as they were deemed less susceptible to be disrupted during the installation of the other devices. Afterward, participants were invited to adjust the height of the stationary bicycle seat. The TV monitor height was then adjusted in order for the screen to be at the participant's eye level. Finally, the Muse headband (Interaxon, Canada) was donned on participants forehead. The experimenter made sure the headband was positioned correctly and was comfortable for the participant.

Participants were then invited to perform a task tutorial (in the form of a PowerPoint presentation). In order to avoid information overload, this tutorial only contained the information about the first video game they were set to play. After the tutorial, participants were invited to practice the first game they were set to play in order to become familiar with the controls. This lasted between 5 and 15 min, depending on participants. Once done, participants completed their first three conditions (the three physical activity levels for the first game). Each of these conditions lasted 10 min. A 2-min baseline was performed before each condition. This baseline consisted in performing the same level of physical activity as the upcoming condition, but without playing any game. Conditions were performed with minimal disruption. The experimenter warned participants who pedaled too slowly or too fast. Additionally, the experimenter tipped players who got stuck for too long in a specific spot. After each condition, participants were invited to complete the two subjective measures questionnaires (NASA-TLX and BORG) and take a short break (roughly 5 min).

Once the three conditions of the first game were completed, participants were presented the tutorial of the second game and performed the remaining three conditions. Two reasons motivated a design in which all conditions of the same game were done subsequently like this. First, we wanted to avoid overloading or confusing participants with shifting game mechanics and controls. Second, we wanted to minimize the lagged effects of stress. Stress is known to influence physiological response even after the stressor is removed (Tassorelli et al., 1995; Qin et al., 2009). While these lagged effects cannot be fully removed from the design, the 5 min breaks between conditions and the non-alternating game conditions helped alleviate this. **Figure 3** shows the experimental setup used. **Figure 4** shows the experimental sequence.

4.8. Physiological Signal Recording

Physiological signals were recorded using the MuSAE Lab EEG Server (MuLES) software (Cassani et al., 2015). MuLES is a LabVIEW software designed to ease simultaneous recording of EEG and other physiological signals. It allows data acquisition of various devices as well as real-time streaming of physiological signal. In our case, data was streamed to a custom-made MATLAB script designed to input markers delimiting the beginning and the end of all experiment trials. Once the data collection finished, a lab assistant manually verified all markers to make sure they correctly matched the experimental trials.

4.9. Signal Processing and Feature Calculation

Physiological signals were loaded in MATLAB using a custom-made parser and trimmed to keep only the relevant parts (baseline and trials). For trials, signals were trimmed into two epochs of 5 min. Baselines were trimmed into epochs of variable length (more or less 2 min).

For the processing of EEG signals, previous works (e.g., Snyder et al., 2015; Bono et al., 2016) have shown that artifact removal methods based on the independent component analysis (ICA) can be successfully employed to enhance EEG data in

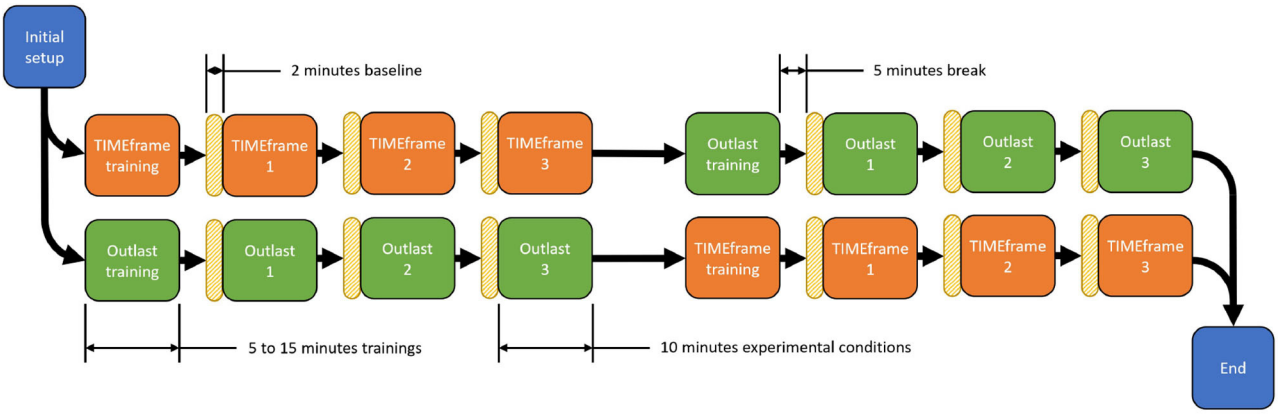


FIGURE 4 | Diagram showing the experimental sequence. After the initial setup, participants completed a training of their first game (5–15 min). They then performed the 3 levels of physical activity for this game. The order of physical activity level was counterbalanced. Each condition lasted 10 min. A 5-min break was inserted between conditions. Before every condition, participants did a 2-min baseline in which they did the same level of physical activity as the upcoming condition, but without playing the game. Half of the participants began with TIMEframe. The other half began with Outlast.

scenarios where artifacts due to physical activity are present. Among these methods, the wavelet-enhanced ICA method, (Castellanos and Makarov, 2006), allows automated artifact removal, and has been proven effective in different scenarios where EEG was acquired with low-density wearable devices (e.g., Cassani et al., 2017; Rosanne et al., 2019). The parameter used for the wICA method in our experiments relied in a threshold $K = 1$ set empirically.

For EEG feature extraction, prefrontal (AF7–AF8) alpha and theta absolute power, and relative gamma power (all locations) were computed since they are known to be associated with stress (Borghini et al., 2014; Minguillon et al., 2016). Prefrontal asymmetry has also been found to be associated with stress (Brouwer et al., 2011) and was computed here between AF8 and AF7. Asymmetry between TP9 and TP10 was also computed for exploratory purposes. Coherence (alpha and beta band) have also been associated with stress in parietal and occipital regions (Giannakakis et al., 2015). As such, it was decided to compute coherence in the closest region available (TP9–TP10) in four frequency subbands (alpha, beta, gamma, and theta). Finally, amplitude modulation features were also computed as per (Falk et al., 2012). Focus is placed here on two specific amplitude modulation features, namely beta modulated by delta (represented as beta-delta) and gamma-delta, given insights reported in Falk et al. (2012), Clerico et al. (2018), and Minguillon et al. (2016). **Table 3** summarizes the EEG features computed, as well as our hypotheses of expected behavior under stress. Expected behavior does not account for possible effects of physical activity.

TABLE 3 | EEG features and their expected behavior under stress.

Type	Feature	Effect
Absolute power	Alpha (AF7)	↓ (Borghini et al., 2014)
	Theta (AF7)	↑ (Borghini et al., 2014)
	Alpha (AF8)	↓ (Borghini et al., 2014)
	Theta (AF8)	↑ (Borghini et al., 2014)
Relative power	Gamma (AF7)	↑ (Minguillon et al., 2016)
	Gamma (AF8)	↑ (Minguillon et al., 2016)
	Gamma (TP9)	↑ (Minguillon et al., 2016)
	Gamma (TP10)	↑ (Minguillon et al., 2016)
Asymmetry	Alpha (AF7-AF8)	↑ (Brouwer et al., 2011)
	Alpha (TP9-TP10)	–
Coherence	Alpha (TP9-TP10)	↑ (Giannakakis et al., 2015)
	Beta (TP9-TP10)	↓ (Giannakakis et al., 2015)
	Gamma (TP9-TP10)	–
	Theta (TP9-TP10)	–
Amplitude modulation	Beta-delta (AF7)	–
	Gamma-delta (AF7)	–
	Beta-delta (AF8)	–
	Gamma-delta (AF8)	–
	Beta-delta (TP9)	–
	Gamma-delta (TP9)	–
	Beta-delta (TP10)	–
	Gamma-delta (TP10)	–

Expected behaviors do not account for possible effects of physical activity.

For the ECG signals, in turn, a variation of the Pan-Tompkins algorithm was used to obtain the interbeat interval time series (Behar et al., 2018). Interbeat intervals were subsequently processed to remove outliers and improbable points. Heart rate variability features, frequently investigated as correlates of stress, were then computed (Castaldo et al., 2015). These features include the heart rate, the standard deviation of interbeat intervals, the power of the high frequency band, and the low-frequency to high-frequency (LF/HF) ratio. Moreover, the breathing signal from the BioHarness 3 was downsampled from 18 to 6 Hz and filtered to remove noise (low-pass, Chebychev, 2 Hz, 8th order). Features previously shown to be modulated by stress were then computed, including breathing

TABLE 4 | Peripheral features and their expected behavior under stress.

Modality	Feature	Effect
Heart	Heart rate	↑ (Castaldo et al., 2015)
	SDNN	↑ (Castaldo et al., 2015)
	HF power	↓ (Castaldo et al., 2015)
	LF/HF	↑ (Castaldo et al., 2015)
Breathing	Breathing rate	↑ (Rainville et al., 2006)
	Variability	↑ (Vlemincx et al., 2013)
	Sigh rate	↑ (Vlemincx et al., 2013)
Electrodermal	Level	↑ (Reinhardt et al., 2012)
	Number of responses	↑ (Reinhardt et al., 2012)
	Rel. LF power	↑ (Posada-Quintero et al., 2016a)
Temperature	Average temperature	↓ (Kreibig et al., 2007)
	Temperature delta	↓ (Kreibig et al., 2007)
Blood volume pulse	Minimum BVP	↑ (Kreibig et al., 2007)
	Maximum BVP	↑ (Kreibig et al., 2007)

Expected behaviors does not account for possible effects of physical activity.

rate (computed by counting the peaks of the filtered signals); breathing variability, computed using sample entropy ($m = 2$, $r = 0.5$) (Vlemincx et al., 2013); and sigh rate, where a sigh is defined as a breath where the amplitude exceeded one standard deviation of the normal breathing amplitude for the condition.

From the E4 wristband, the electrodermal signal was first filtered (low-pass, Chebychev, 1 Hz, 8th order) and the features computed include the electrodermal level (normalized average of the baselines) and the number of electrodermal responses (Boucsein, 2012). In addition, relative low frequency power (0.045–0.15 Hz, LF power) was also computed since recent works suggest that it might be associated with stress (Posada-Quintero et al., 2016a). The E4 wristband was also used to measure temperature. No particular processing was performed on the skin temperature signal. Since stress is known to affect temperature, it was decided to compute the average temperature level and the delta (difference between the end and the initial temperature of a condition) temperature (Kreibig et al., 2007). Moreover, blood volume pulse level was normalized in reference to the average of all baselines of each participant. The minimum and maximum blood volume pulse levels were computed to approximate relative diastolic and systolic pressures. **Table 4** summarizes the peripheral features computed, as well as our hypotheses of expected behavior under stress. As previously, expected behaviors do not account for possible effects of physical activity.

4.10. Database Availability

The PASS database is part of a larger project on operator functional state monitoring aimed at building models that take into account mental workload, stress and physical fatigue. In a related work, we describe the WAUC dataset, which presents an experimental protocol to modulate mental workload and physical activity (Albuquerque et al., submitted). Both datasets are available online for download at http://musaelab.ca/pass-database/. Both the PASS and WAUC databases include raw physiological signals, subjective responses, and additional documentation, such as markers information.

4.11. Modeling

To assess the discriminatory power of the explored features, machine learning models were developed for stress level classification, i.e., classifying between no-stress (TIMEframe) and stress (Outlast) conditions. All physical activity levels were combined in our analyses. This was done in order to see if it was possible to classify stress even if the current level of physical activity is unknown by the classifier. Accounting for missing data, there were 264 samples for TIMEframe and 248 samples for Outlast. Here, a support vector classifier was used (Smets et al., 2019) and two testing schemes were implemented: k-fold and leave-one-participant-out (LOPO).

Both the k-fold and the LOPO scheme used a nested cross-validation scheme. For the k-fold, samples were folded in five-folds for testing. The remaining 4 folds of the samples (for each testing fold) were subdivided again into five-folds to perform the validation. One fifth of these were used for validation. The rest was used for training. For the LOPO scheme, samples were folded per participant for testing. One fifth of the remaining participants (for each testing fold) was used for validation. The rest were used for training. Model hyperparameters (box constraint and lambda) were optimized using Bayesian optimization.

Models are tested using various feature subsets, namely one model per EEG feature subtype (total of five models), one per peripheral feature subtype (five total), one model for all combined EEG features, one model for all combined peripheral features, and, lastly, one model fusing both the EEG and peripheral features. Cohen's kappa is used to gauge classifier performance. Cohen's kappa is a measure that express the agreement between true class labels and models prediction (Billinger et al., 2012). This measure is commonly used in the B/BCI literature (Schlögl et al., 2005; Hasan et al., 2015). A Cohen's kappa of 0 means that the model is doing no better than chance (i.e., the accuracy would be close to 50% if classes were balanced). A Cohen's kappa of 1 means that the model is perfect (i.e., 100% accuracy).

5. DATABASE VALIDATION: EXPERIMENTAL RESULTS

The majority of the participants completed all six experimental conditions. Five participants decided to not perform the Outlast scenario and two participants did not fully complete TIMEframe scenarios. The most common stated cause for early interruption was nausea (possibly induced by the proximity with the screen).

Only one participant reported smoking. No participant reported suffering from hypertension. Subjective weight was reported on a four-point scale (insufficient, normal, excess, great excess). Participants reported having either a normal weight (36 participants) or an excess of weight (11 participants). One participant did not answer the weight question and none

TABLE 5 | Descriptive statistics of subjectives measures.

Variable	Condition average (stress–physical activity)					
	Low-0 km/h	Low-18 km/h	Low-24 km/h	High-0 km/h	High-18 km/h	High-24 km/h
NASA-TLX						
Mental demand	3.8 ± 0.8	4.1 ± 1.0	4.4 ± 0.7	11.1 ± 1.2	11.2 ± 1.1	11.6 ± 1.3
Physical demand	2.4 ± 0.9	5.4 ± 0.8	8.2 ± 1.0	3.1 ± 0.8	7.5 ± 1.1	9.8 ± 1.4
Temporal demand	3.1 ± 0.7	4.0 ± 0.8	4.4 ± 1.0	9.5 ± 1.5	10.0 ± 1.3	11.4 ± 1.4
Performance	11.2 ± 1.7	11.6 ± 1.5	11.2 ± 1.7	12.6 ± 1.7	13.1 ± 1.4	12.3 ± 1.5
Effort	6.8 ± 1.5	7.6 ± 1.3	8.6 ± 1.2	11.1 ± 1.4	11.5 ± 1.1	12.4 ± 1.2
Frustration	4.2 ± 1.3	4.2 ± 1.2	4.1 ± 1.3	9.4 ± 1.6	9.4 ± 1.6	10.0 ± 1.5
Stress*	2.3 ± 0.6	2.5 ± 0.6	2.4 ± 0.6	13.5 ± 1.6	13.0 ± 1.4	13.2 ± 1.4
Fear*	1.1 ± 0.1	1.1 ± 0.1	1.1 ± 0.1	12.2 ± 1.7	11.3 ± 1.5	12.6 ± 1.5
BORG						
After condition	7.0 ± 0.6	8.5 ± 0.6	9.8 ± 0.7	8.1 ± 0.6	10.0 ± 0.7	11.2 ± 0.7
After break	7.0 ± 0.4	8.0 ± 0.6	8.8 ± 0.7	8.1 ± 0.6	8.9 ± 0.7	10.1 ± 0.8

Stress and fear are not part of the original NASA-TLX. See **Table 2.*

reported an insufficient weight or a great excess of weight. A majority of participants reported doing at least 30 min of exercise per day (33 participants, one did not answer). Regarding job activity levels, twenty-five participants reported having a sedentary job (e.g., office job), 13 reported having a low physical job (e.g., housekeeping, woodworking), and only three reported having a moderate physical job (e.g., construction, farming). No participant reported having a heavy-physical job (e.g., carpentry).

Moreover, in the original experimental design, all three scenes used in Outlast (i.e., Admin, Ward, Sewers) were intended to be considered as high stress conditions. It is possible, however, that some scenes were not as stressful as others. To verify this, a preliminary set of repeated measures models were fitted using only the data from Outlast session. This set used the NASA-TLX stress and fear questions as independent variable. Physical activity levels (0, 18, 24 km/h) and condition (Admin, Ward, Sewers) were used as dependant variables. Results of the repeated measure ANOVA suggest that there were no differences between all three Outlast scenes ($p_{stress} > 0.05$, $p_{fear} > 0.05$). Therefore, all Outlast scenes will be pooled under high stress in the subsequent analysis.

The following section will detail the results of the subjective, neurophysiological and peripheral measures, as well as the modeling analysis in order to validate the protocol and database.

5.1. Subjective Results

Table 5 reports the average scores and the mean confidence interval (confidence level of 95%) of the two subjective questionnaires across all six conditions. As can be seen, while the stress levels of the video games had an effect on the reported physical demand scores, the different physical activity levels produced no difference in the reported stress levels scores.

To better understand the effects of stress and physical activity had on subjective measures, a series of repeated measures ANOVAs are performed on NASA-TLX and BORG responses. For dimensions of the subjective rating that did not have normally distributed residuals, we performed a Friedman non-parametric test (**Table 6**). Stress, physical activity level and the interaction between both are used as independent variables. Greenhouse-Geisser correction of the p values was used when assumption of sphericity was violated. The significance level was ($p < 0.005$) after Bonferroni correction was used for multiple comparisons. Results show that the stress manipulation had an effect on most of the subjective variables (except performance) as well as on the two BORG measures. Physical activity had an effect on NASA-TLX physical demand, temporal demand, and effort, as well as on the two BORG measures. No relevant interaction was found.

5.2. EEG Results

Table 7 reports the average values and the mean confidence interval (confidence level of 95%) of the selected EEG features across all six conditions. To analyze these results, repeated measures ANOVA is performed with the same independent variables as for **Table 6**. Greenhouse-Geisser correction of the p values was used when assumption of sphericity was violated. The significance level was ($p < 0.0028$) after Bonferroni correction was used for multiple comparisons. **Table 8** reports the ANOVA results.

As can be seen, prefrontal absolute power of alpha and theta was not significantly altered by the experimental manipulations. The ANOVA did not reveal any effects on relative prefrontal gamma. However, results suggest that relative gamma in temporal-parietal regions was higher during the stress condition (TP9, $p < 0.001$, $\eta_p^2 = 0.30$; TP10, $p < 0.001$, $\eta_p^2 = 0.30$), although they were not affected by physical activity, but an interaction was present at this location ($p = 0.002$, $\eta_p^2 = 0.17$). The asymmetry and coherence features did not reveal any effect or interaction. Amplitude modulation features showed sensitivity to the stress manipulation. This sensitivity appeared higher for gamma modulated by delta for temporal-parietal regions (TP9, $p < 0.001$, $\eta_p^2 = 0.28$; TP10, $p < 0.001$, $\eta_p^2 = 0.28$). Globally,

TABLE 6 | Results of repeated measures ANOVA for subjective measures.

Independent variable	Stress			Physical activity			Stress × Physical activity		
	F	*p*	η_p^2	*F*	*p*	η_p^2	*F*	*p*	η_p^2
NASA-TLX									
Mental demand‡	244.9	< 0.001	0.85	3.6	0.031	0.08	0.3	0.722	0.00
Physical demand†,‡,◇	21.8	< 0.001	0.35	103.4	< 0.001	0.72	3.2	0.046	0.07
Temporal demand‡,◇	106.0	< 0.001	0.72	8.6	< 0.001	0.17	0.6	0.533	0.01
Performance	6.6	0.013	0.14	0.6	0.525	0.01	0.1	0.854	0.00
Effort†,‡,◇	66.8	< 0.001	0.62	8.8	< 0.001	0.18	0.2	0.767	0.00
Frustration	53.9	< 0.001	0.57	0.5	0.557	0.01	0.5	0.564	0.01
Stress*	237.1	< 0.001	0.85	0.1	0.894	0.00	0.0	0.927	0.00
Fear*	209.4	< 0.001	0.83	2.7	0.067	0.06	2.8	0.063	0.06
BORG									
After condition†,‡,◇	36.2	< 0.001	0.47	58.0	< 0.001	0.59	2.3	0.101	0.05
After break†,‡,◇	20.2	< 0.001	0.33	20.0	< 0.001	0.33	0.3	0.700	0.00

Stress and fear are not part of the original NASA-TLX. See **Table 2.*

†*Difference found for multiple comparison test (Tukey-Kramer) between 0 and 18 km/h ($p < 0.05$).*

‡*Difference found for multiple comparison test (Tukey-Kramer) between 0 and 24 km/h ($p < 0.05$).*

◇*Difference found for multiple comparison test (Tukey-Kramer) between 18 and 24 km/h ($p < 0.05$).*

TABLE 7 | Descriptive statistics of EEG features.

Variable (unit)	Condition average (stress–physical activity)					
	Low-0 km/h	Low-18 km/h	Low-24 km/h	High-0 km/h	High-18 km/h	High-24 km/h
Abs. power (dB)						
Alpha (AF7)	0.60 ± 0.38	0.68 ± 0.40	0.67 ± 0.41	0.70 ± 0.44	0.60 ± 0.38	0.52 ± 0.36
Theta (AF7)	1.14 ± 0.71	1.28 ± 0.73	1.28 ± 0.76	1.31 ± 0.82	1.09 ± 0.72	0.99 ± 0.66
Alpha (AF8)	0.87 ± 0.23	1.02 ± 0.28	0.96 ± 0.27	1.04 ± 0.27	0.88 ± 0.28	0.95 ± 0.26
Theta (AF8)	1.66 ± 0.44	1.94 ± 0.52	1.84 ± 0.50	1.97 ± 0.51	1.73 ± 0.53	1.83 ± 0.48
Rel. power (10^{-3})						
Gamma (AF7)	179.4 ± 32.6	178.1 ± 35.7	166.1 ± 33.6	200.4 ± 38.1	161.1 ± 37.2	163.8 ± 38.3
Gamma (AF8)	250.8 ± 38.5	228.6 ± 41.1	221.3 ± 40.6	225.5 ± 34.4	200.0 ± 37.8	222.5 ± 39.7
Gamma (TP9)	56.3 ± 20.9	40.7 ± 10.4	52.0 ± 13.1	79.9 ± 21.0	55.5 ± 13.2	56.3 ± 15.5
Gamma (TP10)	54.5 ± 11.1	38.4 ± 9.7	41.7 ± 10.1	70.0 ± 14.8	50.2 ± 12.5	52.8 ± 13.3
Asymmetry (dB)						
Alpha (AF7-AF8)	0.90 ± 0.45	0.90 ± 0.43	0.88 ± 0.44	0.89 ± 0.48	0.82 ± 0.43	0.90 ± 0.47
Alpha (TP9-TP10)	−0.04 ± 0.12	0.02 ± 0.11	0.04 ± 0.11	0.11 ± 0.19	0.01 ± 0.15	0.00 ± 0.15
Coherence (–)						
Alpha (TP9-TP10)	0.53 ± 0.06	0.53 ± 0.06	0.52 ± 0.05	0.49 ± 0.07	0.49 ± 0.06	0.48 ± 0.07
Beta (TP9-TP10)	0.34 ± 0.05	0.30 ± 0.04	0.28 ± 0.04	0.31 ± 0.06	0.27 ± 0.05	0.27 ± 0.05
Gamma (TP9-TP10)	0.30 ± 0.06	0.25 ± 0.05	0.23 ± 0.04	0.28 ± 0.06	0.23 ± 0.04	0.21 ± 0.04
Theta (TP9-TP10)	0.65 ± 0.06	0.67 ± 0.06	0.66 ± 0.06	0.61 ± 0.07	0.63 ± 0.07	0.60 ± 0.08
AM (10^{-3})						
Beta-delta (AF7)	83.8 ± 10.2	75.6 ± 9.1	77.1 ± 10.9	89.4 ± 10.0	80.2 ± 10.4	76.5 ± 10.4
Gamma-delta (AF7)	81.3 ± 12.2	82.7 ± 14.2	78.3 ± 12.6	91.2 ± 14.7	78.3 ± 14.6	79.1 ± 15.5
Beta-delta (AF8)	74.3 ± 6.2	66.1 ± 5.1	66.9 ± 6.0	79.7 ± 5.1	71.8 ± 5.8	73.9 ± 5.8
Gamma-delta (AF8)	98.9 ± 12.4	90.4 ± 13.4	88.2 ± 13.2	101.2 ± 10.6	82.5 ± 11.9	89.4 ± 12.9
Beta-delta (TP9)	62.4 ± 8.6	54.7 ± 11.5	56.7 ± 12.0	73.1 ± 10.6	60.9 ± 11.5	60.4 ± 11.4
Gamma-delta (TP9)	35.7 ± 9.2	27.9 ± 9.5	32.4 ± 10.5	43.7 ± 11.5	34.8 ± 10.8	34.4 ± 11.8
Beta-delta (TP10)	59.8 ± 8.2	52.1 ± 9.5	52.3 ± 10.4	71.9 ± 9.3	58.2 ± 10.6	59.0 ± 11.0
Gamma-delta (TP10)	31.5 ± 7.6	26.0 ± 8.7	27.7 ± 9.1	39.8 ± 9.3	31.6 ± 9.3	31.9 ± 9.3

TABLE 8 | Results of repeated measures analysis of variance for EEG features.

Independent variable	Stress			Physical activity			Stress × physical activity		
	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2
Abs. power									
Alpha (AF7)	0.4	0.533	0.01	0.9	0.409	0.02	0.8	0.442	0.02
Theta (AF7)	0.3	0.559	0.01	0.8	0.438	0.02	0.7	0.475	0.02
Alpha (AF8)	0.0	0.949	0.00	0.5	0.587	0.01	3.9	0.026	0.10
Theta (AF8)	0.0	0.961	0.00	0.2	0.762	0.01	2.9	0.059	0.08
Rel. power									
Gamma (AF7)	2.1	0.159	0.05	1.3	0.281	0.03	2.8	0.076	0.07
Gamma (AF8)†,‡	0.6	0.802	0.00	5.2	0.009	0.13	2.0	0.139	0.05
Gamma (TP9)	15.1	< 0.000	0.30	2.3	0.130	0.06	7.3	0.002	0.17
Gamma (TP10)†	15.7	< 0.001	0.30	7.0	0.004	0.16	0.7	0.476	0.02
Asymmetry									
Alpha (AF7-AF8)	1.14	0.293	0.03	0.8	0.452	0.02	0.02	0.846	0.00
Alpha (TP9-TP10)	0.6	0.457	0.02	0.2	0.762	0.00	1.0	0.356	0.03
Coherence									
Alpha (TP9-TP10)	4.7	0.037	0.12	0.1	0.932	0.00	0.2	0.844	0.00
Beta (TP9-TP10)†	2.1	0.159	0.05	5.0	0.018	0.12	0.5	0.951	0.00
Gamma (TP9-TP10)†,‡	1.4	0.247	0.04	10.0	< 0.001	0.22	0.9	0.409	0.02
Theta (TP9-TP10)	4.9	0.032	0.12	0.1	0.909	0.00	0.7	0.488	0.02
AM									
Beta-delta (AF7)	2.9	0.099	0.07	2.6	0.096	0.07	2.7	0.082	0.07
Gamma-delta (AF7)	4.1	0.051	0.10	0.6	0.518	0.02	3.1	0.061	0.08
Beta-delta (AF8)†,‡	6.8	0.014	0.16	8.0	< 0.001	0.18	0.1	0.943	0.00
Gamma-delta (AF8)†,‡	0.5	0.482	0.01	5.6	0.008	0.13	1.9	0.149	0.05
Beta-delta (TP9)	6.9	0.012	0.16	2.1	0.097	0.07	1.5	0.222	0.04
Gamma-delta (TP9)	13.8	< 0.001	0.28	1.7	0.196	0.05	5.3	0.008	0.13
Beta-delta (TP10)†	8.9	0.005	0.20	3.6	0.043	0.09	1.4	0.262	0.04
Gamma-delta (TP10)†	14.3	< 0.001	0.28	4.4	0.019	0.11	2.1	0.140	0.05

†*Difference found for multiple comparison test (Tukey-Kramer) between 0 and 18 km/h ($p < 0.05$).*
‡*Difference found for multiple comparison test (Tukey-Kramer) between 0 and 24 km/h ($p < 0.05$).*

amplitude modulation was elevated by stress and lowered by physical activity.

5.3. Peripheral Results

Table 9 reports the average values and the mean confidence interval (confidence level of 95%) of the selected peripheral features across all six conditions. As previously, repeated measures ANOVA is performed with Greenhouse-Geisser correction of the p values was used when assumption of sphericity was violated. The significance level was ($p < 0.0036$) after Bonferroni correction was used for multiple comparisons. **Table 10** reports the ANOVA results.

For the cardiac features, ANOVA suggests that heart rate rose as physical activity was more intense ($p < 0.001, \eta_p^2 = 0.23$). The SDNN was also significantly higher during high stress ($p = 0.002, \eta_p^2 = 0.23$) and decreased with more intense physical activity ($p < 0.001, \eta_p^2 = 0.19$). No effects or interactions were detected for spectral features of heart rate variability (HR, LF/HF ratio). Breathing rate was higher in the stress condition ($p < 0.001, \eta_p^2 = 0.26$) and higher during physical activity, although it was not affect by physical activity. ANOVA revealed an effect of physical activity on sigh rate ($p = 0.002, \eta_p^2 = 0.17$). **Table 9** suggests that sigh rate was higher in presence of physical activity, but slightly higher for the 18 km/h level of physical activity. No effects or interaction was detected for breathing variability. Electrodermal features did not reveal any effect of experimental conditions, although some positive trends could be observed (e.g., number of responses, $p_{stress} = 0.077$, $p_{physical} = 0.071$; Rel. LF power, $p_{stress} = 0.074$, $p_{physical} = 0.091$). Temperature also did not appear to vary across the two experimental manipulations. However, results suggest that the temperature delta was significantly affected by the stress manipulation ($p = 0.003, \eta_p^2 = 0.23$). More specifically, temperature delta was much lower during high stress conditions. ANOVA also suggested that minimum and maximum BVP were much higher during high stress conditions (minimum, $p < 0.001, \eta_p^2 = 0.30$; maximum, $p = 0.003, \eta_p^2 = 0.23$). No effect of physical activity or interaction was found for temperature or BVP features.

TABLE 9 | Descriptive statistics of peripheral measures.

Variable (unit)	Condition average (stress–physical activity)					
	Low-0 km/h	**Low-18 km/h**	**Low-24 km/h**	**High-0 km/h**	**High-18 km/h**	**High-24 km/h**
Cardiac						
Heart rate (*bpm*)	85.0 ± 3.9	91.5 ± 4.0	94.6 ± 4.2	86.7 ± 4.1	92.9 ± 4.0	95.5 ± 4.1
SDNN (*ms*)	46.5 ± 5.6	37.3 ± 4.5	36.3 ± 5.0	53.8 ± 6.3	42.0 ± 4.6	41.0 ± 5.9
HF power (ms^2)	7.1 ± 1.8	9.1 ± 2.7	9.0 ± 2.2	6.1 ± 1.7	6.4 ± 1.5	8.1 ± 2.6
LF/HF (–)	4.3 ± 0.7	3.9 ± 0.8	3.7 ± 0.8	4.4 ± 1.1	4.7 ± 1.1	4.3 ± 1.1
Breathing						
Breathing rate (*breath/min*)	22.4 ± 1.5	26.0 ± 1.6	24.9 ± 1.1	23.4 ± 1.4	27.0 ± 1.5	25.7 ± 1.4
Variability (–)	0.71 ± 0.10	0.83 ± 0.09	0.74 ± 0.10	0.71 ± 0.09	0.81 ± 0.10	0.73 ± 0.11
Sigh rate (min^{-1})	2.36 ± 0.25	3.26 ± 0.38	2.88 ± 0.25	2.23 ± 0.23	2.84 ± 0.34	2.60 ± 0.27
EDA						
Level (–)	0.96 ± 0.22	1.06 ± 0.39	1.29 ± 0.37	1.07 ± 0.24	1.24 ± 0.30	1.10 ± 0.21
Number of peaks (*n*)	46.6 ± 3.0	49.2 ± 2.4	48.8 ± 1.8	47.3 ± 2.4	50.7 ± 1.7	50.6 ± 1.3
Rel. LF power (–)	0.10 ± 0.02	0.15 ± 0.03	0.12 ± 0.03	0.13 ± 0.03	0.16 ± 0.03	0.16 ± 0.03
Skin temperature						
Temperature (°C)	33.8 ± 0.6	33.6 ± 0.5	33.6 ± 0.4	33.6 ± 0.6	33.4 ± 0.5	33.4 ± 0.5
Temperature delta (°C)	0.34 ± 0.11	0.24 ± 0.11	0.22 ± 0.12	0.13 ± 0.10	0.12 ± 0.13	0.12 ± 0.10
Blood volume pulse						
Minimum BVP (–)	0.68 ± 0.13	0.78 ± 0.12	0.82 ± 0.17	0.91 ± 0.19	0.95 ± 0.15	1.01 ± 0.24
Maximum BVP (–)	0.72 ± 0.14	0.86 ± 0.14	0.91 ± 0.17	0.94 ± 0.19	0.88 ± 0.13	1.07 ± 0.22

TABLE 10 | Results of repeated measures analysis of variance for peripheral measures.

Independent variable	Stress			Physical activity			Stress × Physical activity		
	F	***p***	η_p^2	***F***	***p***	η_p^2	***F***	***p***	η_p^2
Cardiac									
Heart rate†,‡	2.7	0.109	0.07	10.5	< 0.001	0.23	0.2	0.831	0.00
SDNN†,‡	10.8	0.002	0.23	8.2	< 0.001	0.19	0.2	0.837	0.00
HF power	2.0	0.172	0.06	2.1	0.133	0.06	1.2	0.308	0.04
LF/HF	1.7	0.198	0.05	0.3	0.693	0.01	0.7	0.470	0.02
Breathing									
Breathing rate†	12.9	< 0.001	0.26	5.6	0.007	0.13	0.3	0.727	0.01
Breathing variability	0.3	0.589	0.01	1.4	0.262	0.04	0.3	0.704	0.01
Sigh rate†	6.0	0.019	0.14	7.2	0.002	0.17	0.5	0.587	0.01
EDA									
EDA Level	0.4	0.522	0.01	0.6	0.479	0.02	2.2	0.130	0.06
EDA responses	3.3	0.077	0.09	2.9	0.071	0.08	0.4	0.639	0.01
EDA Rel. LF power	3.4	0.074	0.09	2.5	0.091	0.07	0.2	0.828	0.01
Skin temperature									
Temperature	3.2	0.082	0.08	1.0	0.359	0.03	0.2	0.817	0.00
Temperature delta	10.4	0.003	0.23	0.7	0.479	0.02	0.7	0.504	0.02
Blood volume pulse									
Minimum BVP	14.2	< 0.001	0.30	0.9	0.402	0.03	0.5	0.568	0.02
Maximum BVP	9.9	0.003	0.23	1.6	0.222	0.04	1.5	0.238	0.04

† *Difference found for multiple comparison test (Tukey-Kramer) between 0 and 18 km/h ($p < 0.05$).*
‡ *Difference found for multiple comparison test (Tukey-Kramer) between 0 and 24 km/h ($p < 0.05$).*

5.4. Modeling Results

A two-way ANOVA is performed using feature subsets and testing schemes as dependent variables. Results from the ANOVA suggest the presence of a significant difference between at least two features subsets ($p < 0.001$), between the two schemes ($p < 0.001$) and an interaction between both factors ($p < 0.001$). To further understand these results, a multiple comparison analysis is performed and Tukey's honest significant difference is used to correct multiple comparisons. **Figure 5** shows the results of these multiple comparisons. For k-fold, the best classification performance was obtained using either all features ($\kappa_{avg} = 0.46$, $acc = 0.73\%$) or all EEG features ($\kappa_{avg} = 0.49$, $acc = 0.74\%$). Amplitude modulation features were the best single type of feature type ($\kappa_{avg} = 0.29$, $acc = 0.65\%$), significantly surpassing the combination of all peripheral features ($\kappa_{avg} = 0.22$, $acc = 0.61\%$), as well as all other single type of features (except relative power features, $\kappa_{avg} = 0.25$, $acc = 0.63\%$). Cardiac features yielded the best performance for single peripheral feature type ($\kappa_{avg} = 0.18$, $acc = 0.60\%$), surpassing EDA, BVP, absolute power and asymmetry features (which all yielded relatively poor results, $\kappa_{avg} < 0.10$, $acc_{avg} = 0.53\%$). Peripheral features provided the most stable results across the two testing schemes. EEG features, on the other hand, all performed very poorly under the LOPO scheme, suggesting that subject-specific models are needed, or more advanced normalization strategies (Albuquerque et al., 2019a).

6. DISCUSSION

As stated in the introduction, this project features two main goals. First, we want to provide a dataset where stress and physical activity are jointly modulated. We also seek to allow exploration of physical activity on artifact generation. Second, we want to provide a dataset that mimics realistic settings to support "in-the-wild" B/BCI development. In the following section, we provide a discussion of the analysis that were performed in order to better characterize the dataset.

6.1. Subjective Analysis

The important effect of the stress manipulation on the custom stress and fear questions suggest that the experimental manipulation was successful. The stress manipulation also had an important effect on mental demand. This result might have been caused by the games design. Despite being similar in terms of game style (first person exploration games), Outlast featured more complex environmental design (e.g., dead ends, hidden passages) than TIMEframe. Outlast also had more complex controls (e.g., using the night mode, running), which might also explain the increased perceived physical demand. Despite knowing that the two games had a predetermined duration (10 min), participants felt that Outlast caused higher temporal demand than TIMEframe. This result might be due to the escape scenes (i.e., escaping from chasing enemies) that were present in all three Outlast scenarios. Overall, it is clear that the stress manipulation caused a high affective stress state, as well as induced some mental stress. This highlights the difficulty in experimentally separating the two forms of stress, a limitation shared with other popular protocols, such as the Trier Social Stress Task (Kudielka et al., 2007).

The physical activity manipulation was also successful, this can be appreciated even with the p values corrected for multiple comparisons between the three level. The descriptive statistics and the straightforward difference between physical activity levels (0, 18, and 24 km/h) suggest that the participants did feel more physical demand as activity levels increased. Results suggest that participants felt a slightly higher temporal demand as physical activity rose. It is important to keep in mind that the higher speed, like any physical activity manipulation, might have induced a higher mental demand on participants. This might have translated into higher subjective temporal demand. Moreover, the effect of physical activity on the effort rating was expected to be higher. However, is it possible that some participants considered this question to concern mental effort, while others physical effort, thus canceling out any potential effects.

6.2. EEG Analysis

Absolute power of alpha and theta did not vary significantly under stress. Since these features were only computed in prefrontal regions, it is possible that they were strongly affected by ocular artifacts. In fact, all prefrontal features computed did not reveal much sensitivity to stress. It could be argued that the task visual load was too high to fully remove all artifacts, suggesting that prefrontal sensitivity to stress could be higher when eyes are closed compared to when eyes are open, as it was reported in Brouwer et al. (2011).

The difference between mental and affective stressors might also explain the absence of effects on prefrontal alpha and theta. In Borghini et al. (2014), the authors mention that the expected decrease of alpha and increase of theta are observed in situation where the task demand is higher. In Giannakakis et al. (2015), authors report several significant differences on absolute power of alpha, beta and theta bands using a more affective than mental stressor (i.e., video segments). Like in Borghini et al. (2014), they do observe lower alpha power in frontal regions (i.e., F3). However, they did not report differences in the locations used in this study. Following this hypothesis, it is also possible that the two video games induced similar mental stress on participants, making it difficult to observe a difference. Additionally, it is also possible that physical activity reduced the experienced stress, thus making it more difficult to be detected. Finally, physical activity might have induced movement on the headset, leading to a poorer contact between the electrodes and the skin. This indeed is a limitation of using the Muse headband.

Notwithstanding, stress had a very clear effect on temporal-parietal relative gamma (on both sides). These results are in line with the hypothesis (see **Table 3**) that relative gamma would rise under stress as per (Minguillon et al., 2016), where the authors focus more on the role of prefrontal relative gamma (rather than temporal-parietal, like in the present study). However, they do report that relative gamma also increased in temporal and parietal regions. In Minguillon et al. (2016), it is suggested that prefrontal relative gamma could be an indicator of mental stress, rather than affective stress (a result supported by previous

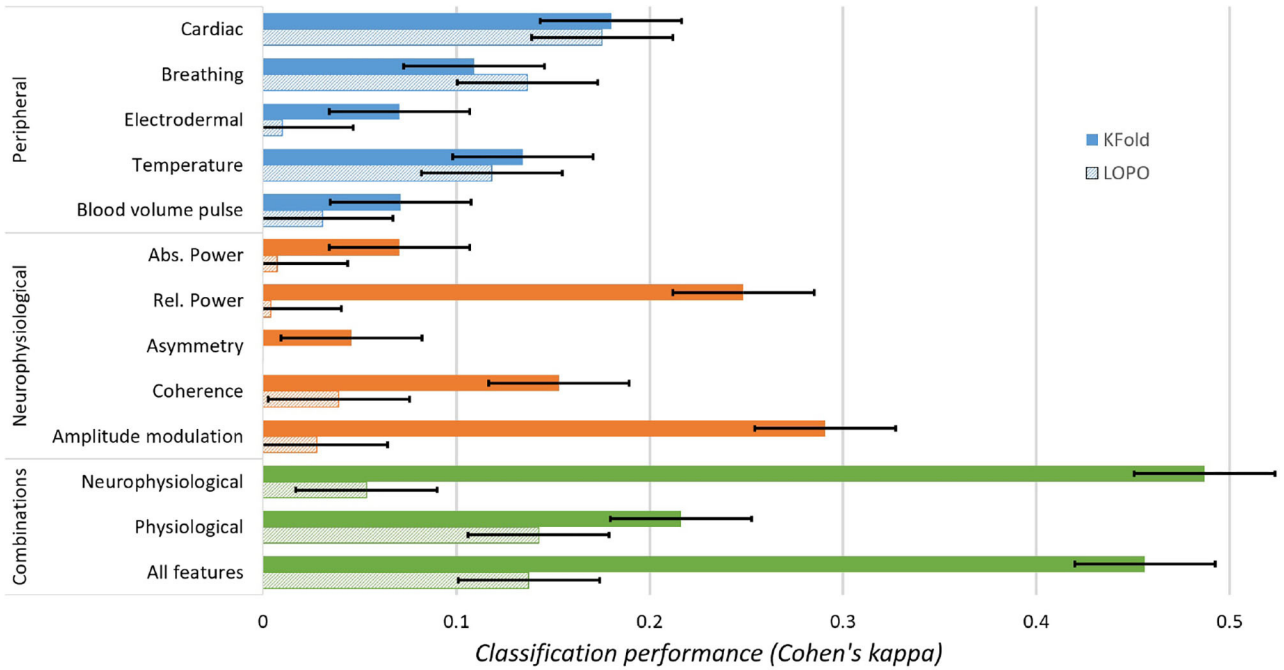


FIGURE 5 | Classification performance (Cohen's kappa) of different feature subsets.

studies, Başar-Eroglu et al., 1996). In another work, temporal and parietal gamma were found to be higher in presence of an affective stressor (Oathes et al., 2008). Given that the current study focused more on affective stress, it is possible that participants experienced similar mental stress in the two games played; this hypothesis is based on the stress level effect on the temporal relative gamma, and the no effect of prefrontal relative gamma. As suggested before, it is possible that ocular and physical activity artifacts seen here prevented detecting a stress effect on prefrontal relative gamma. Moreover, the removal of those artifacts with wICA could have negatively impacted the high frequency components in EEG signals in the prefrontal region (Muthukumaraswamy, 2013; Cassani et al., 2014; Rosanne et al., 2019). Together, these results suggest that stress "in-general" might be associated with the gamma band and that the prefrontal/temporal-parietal predominance might indicate whether this stress is more mental or affective. Further work would be required to confirm this.

Under stress, interhemispheric temporal-parietal coherence (TP9-TP10) was slightly lower, suggesting a less similar neuronal activity between the two regions. This result goes against the hypothesis formulated in **Table 3** (Giannakakis et al., 2015). In Giannakakis et al. (2015), authors found that alpha coherence was higher during stressful video segments compared to relaxed segments (although in parietal region, P3-P4). However, coherence behavior under stress is not well-documented in the literature. In Travis et al. (2010), parietal interhemispheric alpha1 (7.5–10.0 Hz) coherence was higher during meditation compared to control. While we cannot directly compare the TIMEframe game to meditation, it could be argued that the relaxed states enhance interhemispheric coherence.

Lastly, amplitude modulation features yielded several interesting results. Globally, amplitude modulation rose during stress conditions. The apparent larger increase observed on temporal-parietal regions might, once again, have been caused by the hypothesized greater influence of ocular and physical activity artifacts on prefrontal electrodes. Consistent with relative gamma effects, amplitude modulation effects were also greater when observed in the gamma band. It could be argued that the high amplitude modulation observed on gamma (TP9-TP10) are due to fluctuation in experienced stress during the Outlast play session (as opposed to TIMEframe, which induced had a lower and more leveled stress level). In this paper, we explored only a subset of possible amplitude modulation features (i.e., delta-modulated) and future work should explore alternate features. For mental workload assessment, for example, they also showed to be important (Albuquerque et al., 2019b).

6.3. Peripheral Analysis

The increased heart rate observed with physical activity confirms that the physical activity manipulation was effective. As expected, SDNN also rose during high stress conditions (Castaldo et al., 2015). This result reinforces the utility of SDNN as an index of affective stress. However, the observed decrease of SDNN under higher physical activity levels suggest that this feature could have higher predictive power if physical activity of individuals was unknown. Despite being shown sensible to stress in other studies (Kreibig et al., 2007), spectral features of heart rate variability (i.e., HF power, LF/HF ratio) were not significantly affected by the

stress manipulations. Since physical activity is known to change heart rate variability, it is possible that physical activity acted as a confounding factor (Pichon et al., 2004). For example, in Pichon et al. (2004), the LF/HF ratio is reported to decrease as physical activity rises. This behavior might have canceled the expected increase that was hypothesized in **Table 3**. It is also possible that the relatively fast changes in physical activity intensity prevented these features from reaching temporal stability. In addition, the LF/HF ratio has received some criticism as a measure of cognitive and physical aspects of stress, as its correspondence to psychological and physiological states of a person is not unique, and by combining LF and HF one degree of freedom is lost. Future studies could explore the effects on LF and HF separately (von Rosenberg et al., 2017) or investigate potentially more relevant features for ambulant users (e.g., Tiwari et al., 2019, 2020). Finally, it is also possible that spectral features of heart rate variability are more associated with mental stressors than affective stressors. The hypothesis made in **Table 3** are based on Castaldo et al. (2015), which predominantly features mental stressors.

As expected, breathing rate rose under higher physical activity conditions. In concordance with our hypothesis, stress increased the breathing rate (Rainville et al., 2006). The effect size of stress on breathing rate was higher than from physical activity. Surprisingly, sigh rate was lower during stressful conditions, which is opposed to our formulated hypothesis (Vlemincx et al., 2013). Participants might have sighed only once the threat was removed (i.e., after the condition). Given that Outlast's played character is often chased and threatened, it is also possible that participants unconsciously held their breath as not to make noise. The absence of effect on breathing variability might have been caused by the parameters used to compute sample entropy (m, and r). In Vlemincx et al. (2013), authors mentioned that they used $m = 2$ and $r = 0.4$ and these were the parameters used herein. However, it is uncertain if these parameters are optimal for all situations.

The absence of significant effects of stress and physical activity on all electrodermal features was counter-intuitive, as both stress and physical activity have been shown to induce changes in EDA patterns. Placement of the electrodes might partially explain the lack of concordance with the literature. In stress related experiments, electrodes are often placed on the fingers (e.g., Kreibig et al., 2007; Posada-Quintero et al., 2016b, 2018a) or on the foot (e.g., Reinhardt et al., 2012). In setups involving physical activity, it can be more practical to use a wristband (e.g., Gjoreski et al., 2016) as was the case with the current study). It is also possible that the combined affective stressor and physical activity saturated the EDA levels, thus creating a ceiling effect and preventing variability. In Posada-Quintero et al. (2018b), physical activity was manipulated while EDA was recorded. While authors did observe significant difference between the different physical activity levels, they mention that the electrodermal level and the number of responses did not have the sensitivity of spectral features. Physical activity might have also introduced artifacts to the electrodermal measure. Precautions were taken to prevent this: the wrist band was sufficiently tightened to prevent slippage and filtering was applied to the signal to remove higher frequency noise. Finally, it is possible that alternate frequency bands could have achieved improved discriminability. As spectral analysis of EDA is still a relatively undocumented domain, further improvements may be possible.

Lastly, the hypothesis that temperature would be lower during high stress condition (see **Table 4**) was not confirmed. This could be due to the counter-effect of physical activity, which is known to increase body temperature (Lim et al., 2008) even in areas not directly involved in the effort (Chudecka and Lubkowska, 2012). In line with the hypothesis, however, temperature rose much more slowly during stress conditions. As stated in section 2, this could be due to a constriction of the limb arterioles, intended to reduce blood flow in peripheral regions during fight-or-flight situations. This observation matches the increase in blood volume pulse that was induced during stress conditions (Kreibig et al., 2007). It is interesting to note that none of the temperature and blood volume pulse measures were significantly affected by physical activity. Given the short duration of the experimental conditions and the relatively low intensity of the physical task, this behavior is likely not to generalize to all forms of physical activity.

6.4. Modeling Analysis

The goal of the modeling analysis described here was to perform a first validation of the discriminative power of neurophysiological features for stress monitoring under physical activity, and not necessarily to obtain state-of-the-art results (Smets et al., 2019). As such, default classifier parameters were used and classical SVMs were tested. Our ongoing study involves the use of multimodal fusion and classifier optimization to further improve results. The interested reader is referred to Parent et al. (2019b) for more details.

Interestingly, while both EEG and peripheral feature subsets showed similar effect size under stress ($\eta_2^p \approx 0.25$), classification performance differed largely between them. For example, while peripheral features resulted in lower stress prediction performance under the k-fold setting relative to EEG, they generalized better to unseen users in the LOPO scheme. This sensitivity has been reported previously for EEG-based mental workload models (Albuquerque et al., 2019a).

Combining all EEG feature subsets also significantly improved classification results, thus corroborating results reported in the mental workload literature (Albuquerque et al., 2019b). On the other hand, combining EEG with peripheral features did not result in performance gains. Peripheral measures, such as heart rate variability and/or electrodermal activity are often viewed as generic indicators of sympathetic and parasympathetic activation (Billman, 2011; Posada-Quintero et al., 2016a), thus they may provide limited concurrent sources of information, especially in the presence of physical activity.

Moreover, as stated previously, the amplitude modulation features were shown to result in the highest performance under the k-fold setting. Here, only a subset of possible AM features was computed and recent work has suggested that alternate bands can be useful for valence and arousal prediction (Clerico et al., 2018). Future work will explore the full potential of the amplitude modulation features for stress prediction under physical activity.

Within the peripheral modality, cardiac features resulted in the best performance under both testing paradigms. Here, only four cardiac features were explored and relied on time- and frequency-based content. There have been recent innovations in HRV analysis showing that non-linear features may provide improved robustness to noisy data (Tobon et al., 2017), thus improved performance may still be achieved; this is left for future work.

Overall, the modeling analysis results presented herein confirm that affective stressors can induce detectable effects on neuro-physiological signals, despite being in the presence of quickly shifting physical activity. It is hoped that the database provided will allow for other researchers to help advance the knowledge of physiological stress monitoring in the presence of physical activity. This could have important implications for operator functional state monitoring for e.g., first responders.

Lastly, we performed a sanity check to explore the intensity of confound between stress and physical activity. To this end, we performed feature ranking using the recursive feature elimination algorithm. We first found the most important features for stress level detection and trained a classifier on these features to classify physical activity level; we found a Cohen's kappa value of 0.14. In turn, we found the best features for physical activity level classification and used those features to classify stress level; we found a Cohen's kappa of 0.07. Future work could explore the use of physical activity-level aware classification for improved accuracy, as in Sun et al. (2010).

6.5. Future Research Directions

We believe the PASS dataset analysis unlocked many questions and challenges that can be further addressed and investigated by future work. In the following, we summarize some of the research avenues that can be derived from our proposed dataset:

- Design analyses that aim to disentangle the effects of affective and mental stress components on subjective, neurophysiological, and peripheral measures (e.g., evaluate whether different modalities are affected by affective and mental stress in distinct ways);
- Devise EEG artifact removal approaches for data acquired with low-density devices which are also suitable to remove noise generated by physical activity;
- Assess the effect of different EEG referencing approaches on stress detection;
- Explore new features, including (but not limited to), EEG amplitude modulation features that have been linked to mental workload (Albuquerque et al., 2018) or new movement-robust heart rate variability features (Tiwari et al., 2020);
- Develop representation learning pipelines tailored to improve robustness to movement artifacts and inter-subject variability;
- Account for the interplay between stress levels and physical activity by devising stress classification strategies which are conditioned on the current physical activity intensity;
- Explore different state-of-the-art classification schemes and hyperparameter tuning strategies.

7. CONCLUSIONS

The dataset described herein was designed to support the development of physiological stress monitoring models for ambulant users. Two different videogames were used as stress modulators under three physical activity conditions. Our validation results suggest that accurate disambiguation between affective and mental stress effects could be observed even under varying physical activity levels. Validation experiments show features derived from the database to not only corroborate results previously reported in the literature, but to also provide new insights on stress elicitation under physical activity. Lastly, preliminary classification results with popular features and classical classifiers show the promise of stress monitoring of ambulant users with the use of off-the-shelf wearable devices. The collected database, comprised of raw signals, subjective ratings, and triggers, is available for download at http://musaelab.ca/pass-database/.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

Written informed consent was obtained from the individuals for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

All authors: experimental design and writing and reviewing. MP, IA, AT, and RC: statistical analysis and programming. ST and TF: funding and supervision.

ACKNOWLEDGMENTS

The authors acknowledge financial support from NSERC Canada, Thales Canada Inc., MITACS and PROMPT.

REFERENCES

Albuquerque, I., Monteiro, J., Rosanne, O., Tiwari, A., Gagnon, J.-F., and Falk, T. (2019a). "Cross-subject conditional shift estimation for generalized electroencephalography-based mental workload assessment," in *IEEE Systems, Man, and Cybernetics* (Bari). doi: 10.1109/SMC.2019.8914469

Albuquerque, I., Rosanne, O., Gagnon, J.-F., Tremblay, S., and Falk, T. H. (2019b). "Fusion of spectral and spectro-temporal EEG features for mental workload assessment under different levels of physical activity," in *International*

IEEE/EMBS Conference on Neural Engineering (San Francisco, CA: IEEE), 311–314. doi: 10.1109/NER.2019.8717092

Albuquerque, I., Tiwari, A., Gagnon, J.-F., Lafond, D., Parent, M., Tremblay, S., et al. (2018). "On the analysis of EEG features for mental workload assessment during physical activity," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (Miyazaki: IEEE), 538–543. doi: 10.1109/SMC.2018.00101

Albuquerque, I., Tiwari, A., Parent, M., Gagnon, J.-F., Lafond, D., Tremblay, S., et al. (submitted). Wauc: a multi-modal database for mental workload assessment under physical activity. *Front. Neursci. Neural Technol.*

Al-Shargie, F., Kiguchi, M., Badruddin, N., Dass, S. C., Hani, A. F. M., and Tang, T. B. (2016). Mental stress assessment using simultaneous measurement of EEG and fNIRS. *Biomed. Opt. Express* 7, 3882–3898. doi: 10.1364/BOE.7.003882

Aricó, P., Borghini, G., Di Flumeri, G., Colosimo, A., Bonelli, S., Golfetti, A., et al. (2016). Adaptive automation triggered by EEG-based mental workload index: a passive brain-computer interface application in realistic air traffic control environment. *Front. Hum. Neurosci*. 10:539. doi: 10.3389/fnhum.2016.00539

Arsalan, A., Majid, M., Butt, A. R., and Anwar, S. M. (2019). Classification of perceived mental stress using a commercially available EEG headband. *IEEE J. Biomed. Health Inform*. 23, 2257–2264. doi: 10.1109/JBHI.2019.2926407

Banaee, H., Ahmed, M., and Loutfi, A. (2013). Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges. *Sensors* 13, 17472–17500. doi: 10.3390/s131217472

Başar-Eroglu, C., Strüber, D., Schürmann, M., Stadler, M., and Başar, E. (1996). Gamma-band responses in the brain: a short review of psychophysiological correlates and functional significance. *Int. J. Psychophysiol*. 24, 101–112. doi: 10.1016/S0167-8760(96)00051-7

Behar, J. A., Rosenberg, A. A., Weiser-Bitoun, I., Shemla, O., Alexandrovich, A., Konyukhov, E., et al. (2018). Physiozoo: a novel open access platform for heart rate variability analysis of mammalian electrocardiographic data. *Front. Physiol*. 9:1390. doi: 10.3389/fphys.2018.01390

Bernardi, L., Valle, F., Coco, M., Calciati, A., and Sleight, P. (1996). Physical activity influences heart rate variability and very-low-frequency components in Holter electrocardiograms. *Cardiovasc. Res*. 32, 234–237. doi: 10.1016/0008-6363(96)00081-8

Billinger, M., Daly, I., Kaiser, V., Jin, J., Allison, B. Z., Müller-Putz, G. R., et al. (2012). "Is it significant? Guidelines for reporting BCI performance," in *Towards Practical Brain-Computer Interfaces* (Berlin; Heidelberg: Springer), 333–354. doi: 10.1007/978-3-642-29746-5_17

Billman, G. E. (2011). Heart rate variability-a historical perspective. *Front. Physiol*. 2:86. doi: 10.3389/fphys.2011.00086

Bono, V., Das, S., Jamal, W., and Maharatna, K. (2016). Hybrid wavelet and EMD/ICA approach for artifact suppression in pervasive EEG. *J. Neurosci. Methods* 267, 89–107. doi: 10.1016/j.jneumeth.2016.04.006

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., and Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci. Biobehav. Rev*. 44, 58–75. doi: 10.1016/j.neubiorev.2012.10.003

Boucsein, W. (2012). *Electrodermal Activity*. Boston, MA: Springer Science & Business Media. doi: 10.1007/978-1-4614-1126-0

Brouwer, A.-M., Neerincx, M. A., Kallen, V., van der Leer, L., and ten Brinke, M. (2011). EEG alpha asymmetry, heart rate variability and cortisol in response to virtual reality induced stress. *J. Cyberther. Rehabil*. 4, 21–34.

Carroll, D., Turner, J. R., and Rogers, S. (1987). Heart rate and oxygen consumption during mental arithmetic, a video game, and graded static exercise. *Psychophysiology* 24, 112–118. doi: 10.1111/j.1469-8986.1987.tb01870.x

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., et al. (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science* 301, 386–389. doi: 10.1126/science.1083968

Cassani, R., Banville, H., and Falk, T. H. (2015). "Mules: an open source EEG acquisition and streaming server for quick and simple prototyping and recording," in *Proceedings of the 20th International Conference on Intelligent User Interfaces Companion* (Atlanta, GA: ACM), 9–12. doi: 10.1145/2732158.2732193

Cassani, R., Falk, T. H., Fraga, F. J., Cecchi, M., Moore, D. K., and Anghinah, R. (2017). Towards automated electroencephalography-based Alzheimer's disease diagnosis using portable low-density devices. *Biomed. Signal Process. Control* 33, 261–271. doi: 10.1016/j.bspc.2016.12.009

Cassani, R., Falk, T. H., Fraga, F. J., Kanda, P. A. M., and Anghinah, R. (2014). The effects of automated artifact removal algorithms on electroencephalography-based Alzheimer's disease diagnosis. *Front. Aging Neurosci*. 6:55. doi: 10.3389/fnagi.2014.00055

Castaldo, R., Melillo, P., Bracale, U., Caserta, M., Triassi, M., and Pecchia, L. (2015). Acute mental stress assessment via short term HRV analysis in healthy adults: a systematic review with meta-analysis. *Biomed. Signal Process. Control* 18, 370–377. doi: 10.1016/j.bspc.2015.02.012

Castellanos, N. P., and Makarov, V. A. (2006). Recovering EEG brain signals: artifact suppression with wavelet enhanced independent component analysis. *J. Neurosci. Methods* 158, 300–312. doi: 10.1016/j.jneumeth.2006.05.033

Chudecka, M., and Lubkowska, A. (2012). The use of thermal imaging to evaluate body temperature changes of athletes during training and a study on the impact of physiological and morphological factors on skin temperature. *Hum. Movement* 13, 33–39. doi: 10.2478/v10038-012-0002-9

Clerico, A., Tiwari, A., Gupta, R., Jayaraman, S., and Falk, T. H. (2018). Electroencephalography amplitude modulation analysis for automated affective tagging of music video clips. *Front. Comput. Neurosci*. 11:115. doi: 10.3389/fncom.2017.00115

DiDomenico, A., and Nussbaum, M. A. (2008). Interactive effects of physical and mental workload on subjective workload assessment. *Int. J. Indus. Ergon*. 38, 977–983. doi: 10.1016/j.ergon.2008.01.012

Dierolf, A. M., Fechtner, J., Böhnke, R., Wolf, O. T., and Naumann, E. (2017). Influence of acute stress on response inhibition in healthy men: an ERP study. *Psychophysiology* 54, 684–695. doi: 10.1111/psyp.12826

Duncko, R., Johnson, L., Merikangas, K., and Grillon, C. (2009). Working memory performance after acute exposure to the cold pressor stress in healthy volunteers. *Neurobiol. Learn. Mem*. 91, 377–381. doi: 10.1016/j.nlm.2009.01.006

Falk, T. H., Fraga, F. J., Trambaiolli, L., and Anghinah, R. (2012). EEG amplitude modulation analysis for semi-automated diagnosis of Alzheimer's disease. *EURASIP J. Adv. Signal Process*. 2012:192. doi: 10.1186/1687-6180-2012-192

Falk, T. H., Maier, M., et al. (2016). MS-QI: a modulation spectrum-based ECG quality index for telehealth applications. *IEEE Trans. Biomed. Eng*. 63, 1613–1622. doi: 10.1109/TBME.2014.2355135

Feng, S., Tang, M., Quivira, F., Dyson, T., Cuckov, F., and Schirner, G. (2016). "EEGU2: an embedded device for brain/body signal acquisition and processing," in *Proceedings of the 27th International Symposium on Rapid System Prototyping: Shortening the Path from Specification to Prototype* (Pittsburgh, PA), 19–25. doi: 10.1145/2990299.2990304

Fortin-Côté, A., Chamberland, C., Parent, M., Tremblay, S., Jackson, P., Beaudoin-Gagnon, N., et al. (2018). "Predicting video game players–fun from physiological and behavioural data," in *Future of Information and Communication Conference* (Singapore: Springer), 479–495. doi: 10.1007/978-3-030-03402-3_33

Frankenhaeuser, M. (1986). "A psychobiological framework for research on human stress and coping," in *Dynamics of Stress*, M. H. Appley and R. Trumbull (Boston, MA: Springer), 101–116. doi: 10.1007/978-1-4684-5122-1_6

Gagnon, J.-F., Gagnon, O., Lafond, D., Parent, M., and Tremblay, S. (2016). "A systematic assessment of operational metrics for modeling operator functional state," in *PhyCS* (Lisbon: SciTePress), 15–23. doi: 10.5220/0005921600150023

Gaillard, A., and Wientjes, C. (1994). Mental load and work stress as two types of energy mobilization. *Work Stress* 8, 141–152. doi: 10.1080/02678379408259986

Giannakakis, G., Grigoriadis, D., and Tsiknakis, M. (2015). "Detection of stress/anxiety state from EEG features during video watching," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Milan: IEEE), 6034–6037. doi: 10.1109/EMBC.2015.7319767

Gjoreski, M., Gjoreski, H., Luštrek, M., and Gams, M. (2016). "Continuous stress detection using a wrist device: in laboratory and real life," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct* (Heidelberg: ACM), 1185–1193. doi: 10.1145/2968219.2968306

Guitard, T., Bouchard, S., Bernier, F., Dumoulin, S., and Laforest, M. (2010). "The potential of stress indiced with 3-D videogames for teaching stress management

skills to soldiers," in *Proceedings 44th Annual Conference of the Association for Behavioral and Cognitive Therapy* (San Francisco, CA), 18–21.

Gwin, J. T., Gramann, K., Makeig, S., and Ferris, D. P. (2010). Removal of movement artifact from high-density EEG recorded during walking and running. *J. Neurophysiol.* 103, 3526–3534. doi: 10.1152/jn.00105.2010

Harrivel, A. R., Stephens, C. L., Milletich, R. J., Heinich, C. M., Last, M. C., Napoli, N. J., et al. (2017). "Prediction of cognitive states during flight simulation using multimodal psychophysiological sensing," in *AIAA Information Systems-AIAA Infotech@ Aerospace* (Grapevine, TX), 1135. doi: 10.2514/6.2017-1135

Hart, S. G. (2006). "Nasa-task load index (NASA-TLX); 20 years later," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Los Angeles, CA: Sage publications), 904–908. doi: 10.1177/154193120605000909

Hasan, M. R., Ibrahimy, M. I., Motakabber, S., and Shahid, S. (2015). "Classification of multichannel EEG signal by linear discriminant analysis," in *Progress in Systems Engineering*, eds H. Selvaraj, D. Zydek, and G. Chmaj (Cham: Springer), 279–282. doi: 10.1007/978-3-319-08422-0_42

Hjortskov, N., Rissén, D., Blangsted, A. K., Fallentin, N., Lundberg, U., and Søgaard, K. (2004). The effect of mental stress on heart rate variability and blood pressure during computer work. *Eur. J. Appl. Physiol.* 92, 84–89. doi: 10.1007/s00421-004-1055-z

Homma, I., and Masaoka, Y. (2008). Breathing rhythms and emotions. *Exp. Physiol.* 93, 1011–1021. doi: 10.1113/expphysiol.2008.042424

Hovsepian, K., al'Absi, M., Ertin, E., Kamarck, T., Nakajima, M., and Kumar, S. (2015). "cStress: towards a gold standard for continuous stress assessment in the mobile environment," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka: ACM), 493–504. doi: 10.1145/2750858.2807526

Hwang, S., Jebelli, H., Choi, B., Choi, M., and Lee, S. (2018). Measuring workers' emotional state during construction tasks using wearable EEG. *J. Construct. Eng. Manage.* 144:04018050. doi: 10.1061/(ASCE)CO.1943-7862.0001506

Jamal, M. (2007). Job stress and job performance controversy revisited: an empirical examination in two countries. *Int. J. Stress Manage.* 14:175. doi: 10.1037/1072-5245.14.2.175

Jones, M. K., Latreille, P. L., and Sloane, P. J. (2016). Job anxiety, work-related psychological illness and workplace performance. *Brit. J. Indus. Relat.* 54, 742–767. doi: 10.1111/bjir.12159

Kneller, W., Higham, P., and Hobbs, M. (2012). Measuring manual dexterity and anxiety in divers using a novel task at 35-41 m. *Aviat. Space Environ. Med.* 83, 54–57. doi: 10.3357/ASEM.3123.2012

Kreibig, S. D., Wilhelm, F. H., Roth, W. T., and Gross, J. J. (2007). Cardiovascular, electrodermal, and respiratory response patterns to fear-and sadness-inducing films. *Psychophysiology* 44, 787–806. doi: 10.1111/j.1469-8986.2007.00550.x

Kudielka, B. M., Hellhammer, D. H., Kirschbaum, C., Harmon-Jones, E., and Winkielman, P. (2007). Ten years of research with the trier social stress test-revisited. *Soc. Neurosci.* 56:83. doi: 10.1016/B978-012373947-6.00681-4

Lazarus, R. S., and Folkman, S. (1984). *Stress, Appraisal, and Coping*. New York, NY: Springer Publishing Company.

Lim, C. L., Byrne, C., and Lee, J. K. (2008). Human thermoregulation and measurement of body temperature in exercise and clinical settings. *Ann. Acad. Med. Singapore* 37:347.

Marazziti, D., Di Muro, A., and Castrogiovanni, P. (1992). Psychological stress and body temperature changes in humans. *Physiol. Behav.* 52, 393–395. doi: 10.1016/0031-9384(92)90290-I

Matthews, G., and Campbell, S. E. (2010). Dynamic relationships between stress states and working memory. *Cogn. Emot.* 24, 357–373. doi: 10.1080/02699930903378719

Matthews, G., Reinerman-Jones, L., Wohleber, R., Lin, J., Mercado, J., and Abich, J. (2015). "Workload is multidimensional, not unitary: what now?" in *International Conference on Augmented Cognition* (Springer), 44–55. doi: 10.1007/978-3-319-20816-9_5

McRae, A. L., Saladin, M. E., Brady, K. T., Upadhyaya, H., Back, S. E., and Timmerman, M. A. (2006). Stress reactivity: biological and subjective responses to the cold pressor and trier social stressors. *Hum. Psychopharmacol. Clin. Exp.* 21, 377–385. doi: 10.1002/hup.778

Minguillon, J., Lopez-Gordo, M. A., and Pelayo, F. (2016). Stress assessment by prefrontal relative gamma. *Front. Comput. Neurosci.* 10:101. doi: 10.3389/fncom.2016.00101

Muthukumaraswamy, S. (2013). High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations. *Front. Hum. Neurosci.* 7:138. doi: 10.3389/fnhum.2013.00138

Nakamura, K. (2011). Central circuitries for body temperature regulation and fever. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 301, R1207–R1228. doi: 10.1152/ajpregu.00109.2011

Neto, C. A., Pellegrini, L. F., Ferreira, M. S., de Oliveira, S. Jr, and Yanagihara, J. I. (2010). Exergy analysis of human respiration under physical activity. *Int. J. Appl. Thermodyn.* 13, 105–109.

Oathes, D. J., Ray, W. J., Yamasaki, A. S., Borkovec, T. D., Castonguay, L. G., Newman, M. G., et al. (2008). Worry, generalized anxiety disorder, and emotion: evidence from the EEG gamma band. *Biol. Psychol.* 79, 165–170. doi: 10.1016/j.biopsycho.2008.04.005

Ollander, S., Godin, C., Campagne, A., and Charbonnier, S. (2016). "A comparison of wearable and stationary sensors for stress detection," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (Budapest: IEEE), 004362–004366. doi: 10.1109/SMC.2016.7844917

Parent, M., Peysakhovich, V., Mandrick, K., Tremblay, S., and Causse, M. (2019a). The diagnosticity of psychophysiological signatures: can we disentangle mental workload from acute stress with ECG and fNIRS? *Int. J. Psychophysiol.* 146, 139–147. doi: 10.1016/j.ijpsycho.2019.09.005

Parent, M., Tiwari, A., Albuquerque, I., Gagnon, J.-F., Lafond, D., Tremblay, S., et al. (2019b). "A multimodal approach to improve the robustness of physiological stress prediction during physical activity," in *IEEE Systems, Man, and Cybernetics* (Bari). doi: 10.1109/SMC.2019.8914254

Parnandi, A., and Gutierrez-Osuna, R. (2015). A comparative study of game mechanics and control laws for an adaptive physiological game. *J. Multimodal User Interfaces* 9, 31–42. doi: 10.1007/s12193-014-0159-y

Pedersen, B. K., and Saltin, B. (2015). Exercise as medicine-evidence for prescribing exercise as therapy in 26 different chronic diseases. *Scand. J. Med. Sci. Sports* 25, 1–72. doi: 10.1111/sms.12581

Pfaff, M. S. (2012). Negative affect reduces team awareness: the effects of mood and stress on computer-mediated team communication. *Hum. Factors* 54, 560–571. doi: 10.1177/0018720811432307

Pichon, A. P., de Bisschop, C., Roulaud, M., Denjean, A., and Papelier, Y. (2004). Spectral analysis of heart rate variability during exercise in trained subjects. *Med. Sci. Sports Exerc.* 36, 1702–1708. doi: 10.1249/01.MSS.0000142403.93205.35

Porter, A. M., and Goolkasian, P. (2019). Video games and stress: how stress appraisals and game content affect cardiovascular and emotion outcomes. *Front. Psychol.* 10:967. doi: 10.3389/fpsyg.2019.00967

Posada-Quintero, H. F., Florian, J. P., Orjuela-Cañón, A. D., Aljama-Corrales, T., Charleston-Villalobos, S., and Chon, K. H. (2016a). Power spectral density analysis of electrodermal activity for sympathetic function assessment. *Ann. Biomed. Eng.* 44, 3124–3135. doi: 10.1007/s10439-016-1606-6

Posada-Quintero, H. F., Florian, J. P., Orjuela-Cañón, A. D., and Chon, K. H. (2016b). Highly sensitive index of sympathetic activity based on time-frequency spectral analysis of electrodermal activity. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 311, R582–R591. doi: 10.1152/ajpregu.00180.2016

Posada-Quintero, H. F., Florian, J. P., Orjuela-Cañón, A. D., and Chon, K. H. (2018a). Electrodermal activity is sensitive to cognitive stress under water. *Front. Physiol.* 8:1128. doi: 10.3389/fphys.2017.01128

Posada-Quintero, H. F., Reljin, N., Mills, C., Mills, I., Florian, J. P., VanHeest, J. L., et al. (2018b). Time-varying analysis of electrodermal activity during exercise. *PLoS ONE* 13:e0198328. doi: 10.1371/journal.pone.0198328

Qin, S., Hermans, E. J., van Marle, H. J., Luo, J., and Fernández, G. (2009). Acute psychological stress reduces working memory-related activity in the dorsolateral prefrontal cortex. *Biol. Psychiatry* 66, 25–32. doi: 10.1016/j.biopsych.2009.03.006

Quaedflieg, C., Meyer, T., Smulders, F., and Smeets, T. (2015). The functional role of individual-alpha based frontal asymmetry in stress responding. *Biol. Psychology* 104, 75–81. doi: 10.1016/j.biopsycho.2014.11.014

Rainville, P., Bechara, A., Naqvi, N., and Damasio, A. R. (2006). Basic emotions are associated with distinct patterns of cardiorespiratory activity. *Int. J. Psychophysiol.* 61, 5–18. doi: 10.1016/j.ijpsycho.2005.10.024

Random Seed Games (2015). *Timeframe*. Random Seed Games.

Red Barrels Games (2013). *Outlast*. Red Barrels Games.

Reinhardt, T., Schmahl, C., Wüst, S., and Bohus, M. (2012). Salivary cortisol, heart rate, electrodermal activity and subjective stress responses to the Mannheim multicomponent stress test (MMST). *Psychiatry Res*. 198, 106–111. doi: 10.1016/j.psychres.2011.12.009

Righi, S., Mecacci, L., and Viggiano, M. P. (2009). Anxiety, cognitive self-evaluation and performance: ERP correlates. *J. Anxiety Disord*. 23, 1132–1138. doi: 10.1016/j.janxdis.2009.07.018

Rosanne, O., Albuquerque, I., Gagnon, J., Tremblay, S., and Falk, T. H. (2019). "Performance comparison of automated EEG enhancement algorithms for mental workload assessment of ambulant users," in *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)* (San Francisco, CA), 61–64. doi: 10.1109/NER.2019.8716977

Salmon, P. (2001). Effects of physical exercise on anxiety, depression, and sensitivity to stress: a unifying theory. *Clin. Psychol. Rev*. 21, 33–61. doi: 10.1016/S0272-7358(99)00032-X

Schlögl, A., Lee, F., Bischof, H., and Pfurtscheller, G. (2005). Characterization of four-class motor imagery EEG data for the BCI-competition 2005. *J. Neural Eng*. 2:L14. doi: 10.1088/1741-2560/2/4/L02

Schneiderman, N., Ironson, G., and Siegel, S. D. (2005). Stress and health: psychological, behavioral, and biological determinants. *Annu. Rev. Clin. Psychol*. 1, 607–628. doi: 10.1146/annurev.clinpsy.1.102803.144141

Schubert, C., Lambertz, M., Nelesen, R., Bardwell, W., Choi, J.-B., and Dimsdale, J. (2009). Effects of stress on heart rate complexity–a comparison between short-term and chronic stress. *Biol. Psychol*. 80, 325–332. doi: 10.1016/j.biopsycho.2008.11.005

Selye, H. (1985). The nature of stress. *Basal Facts* 7, 3–11. doi: 10.1007/978-1-4757-9573-8_2

Skirbekk, B., Hansen, B. H., Oerbeck, B., Wentzel-Larsen, T., and Kristensen, H. (2012). Motor impairment in children with anxiety disorders. *Psychiatry Res*. 198, 135–139. doi: 10.1016/j.psychres.2011.12.008

Smets, E., De Raedt, W., and Van Hoof, C. (2019). Into the wild: the challenges of physiological stress detection in laboratory and ambulatory settings. *IEEE J. Biomed. Health Inform*. 23, 463–473. doi: 10.1109/JBHI.2018.2883751

Snyder, K. L., Kline, J. E., Huang, H. J., and Ferris, D. P. (2015). Independent component analysis of gait-related movement artifact recorded using EEG electrodes during treadmill walking. *Front. Hum. Neurosci*. 9:639. doi: 10.3389/fnhum.2015.00639

Spielberger, C. D. (2010). "State-trait anxiety inventory," in *The Corsini Encyclopedia of Psychology*, eds C. Nemeroff and W. E. Craighead (American Cancer Society). doi: 10.1002/9780470479216.corpsy0943

Sun, F.-T., Kuo, C., Cheng, H.-T., Buthpitiya, S., Collins, P., and Griss, M. (2010). "Activity-aware mental stress detection using physiological sensors, in *International Conference on Mobile Computing, Applications, and Services* (Santa Clara, CA: Springer), 282–301. doi: 10.1007/978-3-642-29336-8_16

Swan, A. J., Kendall, P. C., Olino, T., Ginsburg, G., Keeton, C., Compton, S., et al. (2018). Results from the child/adolescent anxiety multimodal longitudinal study (camels): functional outcomes. *J. Consult. Clin. Psychol*. 86:738. doi: 10.1037/ccp0000334

Tan, D., and Nijholt, A. (2010). "Brain-computer interfaces and human-computer interaction," in *Brain-Computer Interfaces* (Springer), 3–19. doi: 10.1007/978-1-84996-272-8_1

Tassorelli, C., Micieli, G., Osipova, V., Rossi, F., and Nappi, G. (1995). Pupillary and cardiovascular responses to the cold-pressor test. *J. Auton. Nerv. Syst*. 55, 45–49. doi: 10.1016/0165-1838(95)00026-T

Tiwari, A., Albuquerque, I., Parent, M., Gagnon, J.-F., Lafond, D., Tremblay, S., et al. (2019). Multi-scale heart beat entropy measures for mental workload assessment of ambulant users. *Entropy* 21:783. doi: 10.3390/e21080783

Tiwari, A., Cassani, R., Gagnon, J.-F., Lafond, D., Tremblay, S., and Falk, T. H. (2020). "Prediction of stress and mental workload during police academy training using ultra-short-term heart rate variability and breathing analysis," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (Montreal, QC: IEEE), 4530–4533. doi: 10.1109/EMBC44109.2020.9175414

Tobon, D. P., Jayaraman, S., and Falk, T. H. (2017). Spectro-temporal electrocardiogram analysis for noise-robust heart rate and heart rate variability measurement. *IEEE J. Transl. Eng. Health Med*. 5, 1–11. doi: 10.1109/JTEHM.2017.2767603

Travis, F., Haaga, D. A., Hagelin, J., Tanner, M., Arenander, A., Nidich, S., et al. (2010). A self-referential default brain state: patterns of coherence, power, and eloreta sources during eyes-closed rest and transcendental meditation practice. *Cogn. Process*. 11, 21–30. doi: 10.1007/s10339-009-0343-2

Vachiratamporn, V., Legaspi, R., Moriyama, K., and Numao, M. (2013). "Towards the design of affective survival horror games: an investigation on player affect," in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction* (IEEE), 576–581. doi: 10.1109/ACII.2013.101

Vianna, D. M., and Carrive, P. (2005). Changes in cutaneous and body temperature during and after conditioned fear to context in the rat. *Eur. J. Neurosci*. 21, 2505–2512. doi: 10.1111/j.1460-9568.2005.04073.x

Visnovcova, Z., Mestanik, M., Javorka, M., Mokra, D., Gala, M., Jurko, A., et al. (2014). Complexity and time asymmetry of heart rate variability are altered in acute mental stress. *Physiol. Meas*. 35:1319. doi: 10.1088/0967-3334/35/7/1319

Vlemincx, E., Abelson, J. L., Lehrer, P. M., Davenport, P. W., Van Diest, I., and Van den Bergh, O. (2013). Respiratory variability and sighing: a psychophysiological reset model. *Biol. Psychol*. 93, 24–32. doi: 10.1016/j.biopsycho.2012.12.001

von Rosenberg, W., Chanwimalueang, T., Adjei, T., Jaffer, U., Goverdovsky, V., and Mandic, D. P. (2017). Resolving ambiguities in the LF/HF ratio: LF-HF scatter plots for the categorization of mental and physical stress from HRV. *Front. Physiol*. 8:360. doi: 10.3389/fphys.2017.00360

Wolf, O. T. (2009). Stress and memory in humans: twelve years of progress? *Brain Res*. 1293, 142–154. doi: 10.1016/j.brainres.2009.04.013

Wolpaw, J. R., Birbaumer, N., Heetderks, W. J., McFarland, D. J., Peckham, P. H., Schalk, G., et al. (2000). Brain-computer interface technology: a review of the first international meeting. *IEEE Trans. Rehabil. Eng*. 8, 164–173. doi: 10.1109/TRE.2000.847807

Young, M. S., Brookhuis, K. A., Wickens, C. D., and Hancock, P. A. (2015). State of science: mental workload in ergonomics. *Ergonomics* 58, 1–17. doi: 10.1080/00140139.2014.956151

Conflict of Interest: J-FG and DL were employed by the company Thales Research and Technology Canada.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 18 December 2020
doi: 10.3389/fnins.2020.601402

Prefrontal Asymmetry BCI Neurofeedback Datasets

Fred Charles[1]*, *Caio De Castro Martins*[2] *and Marc Cavazza*[2]*

[1] Faculty of Science and Technology, Bournemouth University, Poole, United Kingdom, [2] School of Computing and Mathematical Sciences, University of Greenwich, London, United Kingdom

Prefrontal cortex (PFC) asymmetry is an important marker in affective neuroscience and has attracted significant interest, having been associated with studies of motivation, eating behavior, empathy, risk propensity, and clinical depression. The data presented in this paper are the result of three different experiments using PFC asymmetry neurofeedback (NF) as a Brain-Computer Interface (BCI) paradigm, rather than a therapeutic mechanism aiming at long-term effects, using functional near-infrared spectroscopy (fNIRS) which is known to be particularly well-suited to the study of PFC asymmetry and is less sensitive to artifacts. From an experimental perspective the BCI context brings more emphasis on individual subjects' baselines, successful and sustained activation during epochs, and minimal training. The subject pool is also drawn from the general population, with less bias toward specific behavioral patterns, and no inclusion of any patient data. We accompany our datasets with a detailed description of data formats, experiment and protocol designs, as well as analysis of the individualized metrics for definitions of success scores based on baseline thresholds as well as reference tasks. The work presented in this paper is the result of several experiments in the domain of BCI where participants are interacting with continuous visual feedback following a real-time NF paradigm, arising from our long-standing research in the field of affective computing. We offer the community access to our fNIRS datasets from these experiments. We specifically provide data drawn from our empirical studies in the field of affective interactions with computer-generated narratives as well as interfacing with algorithms, such as heuristic search, which all provide a mechanism to improve the ability of the participants to engage in active BCI due to their realistic visual feedback. Beyond providing details of the methodologies used where participants received real-time NF of left-asymmetric increase in activation in their dorsolateral prefrontal cortex (DLPFC), we re-establish the need for carefully designing protocols to ensure the benefits of NF paradigm in BCI are enhanced by the ability of the real-time visual feedback to adapt to the individual responses of the participants. Individualized feedback is paramount to the success of NF in BCIs.

Keywords: functional near infrared spectroscopy (fNIRS), PFC asymmetry, visual feedback (VF), neurofeedback (NF), dataset

OPEN ACCESS

Edited by:
Davide Valeriani,
Massachusetts Eye & Ear Infirmary,
Harvard Medical School,
United States

Reviewed by:
Junjie Bu,
University of Science and Technology of China, China
Arnaud Saj,
Université de Genève, Switzerland

***Correspondence:**
Fred Charles
fcharles@bournemouth.ac.uk
Marc Cavazza
m.cavazza@greenwich.ac.uk

Specialty section:
This article was submitted to Neural Technology, a section of the journal Frontiers in Neuroscience

Received: *31 August 2020*
Accepted: *25 November 2020*
Published: *18 December 2020*

Citation:
Charles F, De Castro Martins C and Cavazza M (2020) Prefrontal Asymmetry BCI Neurofeedback Datasets. Front. Neurosci. 14:601402. doi: 10.3389/fnins.2020.601402

1. INTRODUCTION AND RATIONALE

There is growing interest in sharing datasets for Brain-Computer Interfaces (BCI), to facilitate comparison of technical approaches. Their availability is of particular relevance for applications in which there is significant diversity of practice and lack of standardized protocols, such as Neurofeedback (NF) (Ros et al., 2020).

Such datasets make it possible to explore and compare signal acquisition and dynamics, baselining, thresholding, and categorization: this has the potential to identify experimental difficulties and best practice, beyond reproducibility issues.

Among the various neural signals that support BCI, the availability of fNIRS dataset remains scarce (Bak et al., 2019), despite its growing popularity, both for BCI (Naseer and Hong, 2015) and NF applications (Kohl et al., 2020).

In this paper, we introduce three datasets obtained as part of fNIRS BCI experiments. The originality of these datasets is that they were produced in a BCI context yet using a NF paradigm, in which users control their Prefrontal Cortex (PFC) asymmetry. The use of a NF approach to BCI is characterized by an emphasis on RoI activation over long-term effects, often with minimal training compared to clinical uses of NF. It is of particular interest when the RoI signal is not under direct volitional control, as the NF channel assists the user in controlling the signal. Moreover, the feedback channel can be embedded in the interface design itself for added realism. Frontal asymmetry is an important brain signal which has a long history in BCI, for the measure of valence, approach or cognitive workload, and NF. Since frontal signals are of the main elements of fNIRS, this dataset has validity beyond the specific context it has been produced in, which is PFC asymmetry NF.

After a reminder of key concepts in NF, which includes a short discussion of current thinking in fNIRS NF, we discuss the potential interest of our datasets to the wider fNIRS and NF community, and describe several data formats supported by our dataset to facilitate processing by various software packages and data-oriented programming languages. In the remainder of the text we will refer to our three datasets as follow:

- **ANG** (Aranyi et al., 2015b) is derived from an anger-expressing BCI experiment.
- **RAP** (Aranyi et al., 2016) investigates rapport with a virtual character endowed with full facial expressions.
- **HEU** (Cavazza et al., 2017) uses BCI input to a hybrid human-AI system.

1.1. PFC Asymmetry in Neuroscience Research

One of the major challenges for BCI is to relate neural signals to specific cognitive processes, or to an element of user experience. For affective BCI, this is rendered even more difficult by the weakness of locationist hypotheses (Lindquist et al., 2012). However, there is substantial evidence linking prefrontal cortex (PFC) asymmetry to the approach/withdrawal dimension (Davidson, 2004): the paradox being that an area associated to high-level integrative cognitive processes is also the locus of a rather basic dimension. This dimension has been shown to underpin higher-level behavioral elements including motivation, risk-taking, aggression, and empathy. Moreover, it has been associated to clinical conditions, such as addiction, eating disorders, gambling and depression. While PFC asymmetry has been primarily associated with approach/withdrawal it has also been shown to be highly correlated with valence, as well as cognitive workload.

Historically, interest in PFC asymmetry has stemmed from research in affective and social neuroscience. Another significant use of PFC asymmetry has been early NF experiments, primarily for the treatment of depression (Rosenfeld et al., 1995). PFC asymmetry has been later adopted as a BCI technology taking advantage of the above results, and has been used for affective computing (Mühl and Heylen, 2009) cognitive workload measurement (Fishburn et al., 2014; Peck et al., 2014; Barth et al., 2016; Maior et al., 2020) or assessment of aesthetic response (Karran et al., 2015; Cartocci et al., 2016).

Most of early work on PFC NF has taken place using EEG signals. There are several reasons for that: despite the lack of spatial resolution, it is still possible to capture a meaningful PFC asymmetry signal from *F3* and *F4* electrodes. The existence of a stable PFC asymmetry EEG baseline in the alpha spectrum and the trait and state properties of the signal (Coan and Allen, 2002) facilitates the design of PFC NF experiments. With the increasing availability of fNIRS equipment, it appeared as an interesting alternative to EEG with less sensitivity to a range of artifacts and increased specificity and spatial resolution considering that the RoI is close to the surface hence easily accessible to infrared sensors. Sitaram et al. (2009) were amongst the first to suggest that signals based on metabolic activity could be equally suited to BCI than electrical signals. fNIRS has become the method of choice as DLPFC is readily accessible via lateral optodes (Ernst et al., 2013). It has been used as a measurement tool (Hirshfield et al., 2007) and as a BCI (Solovey et al., 2009; Afergan et al., 2014; Naseer and Hong, 2015; Hong et al., 2020) to support NF research and even clinical experiments.

1.2. Neurofeedback Concepts

Our dataset has been entirely produced through NF BCI experiments and it is worth summarizing some of the main concepts attached to NF technical implementations and subject behavior. In naïve terms, NF consists in facilitating the activation of a region of interest in the subject's brain through the real-time display of a feedback signal that represents how successful they are in activating that region[1]. NF is generally considered an operant conditioning mechanism, and subjects tend to develop or improve this ability through training, although this ability shows great individual variability, some subjects demonstrating it from the very first testing sessions while others hardly develop it, a phenomenon close to BCI illiteracy (Lee et al., 2014; Trambaiolli et al., 2018).

[1]We are primarily considering here NF systems for which there is an element of regional specificity or spatial resolution, which is generally the case for fNIRS and fMRI-based NF.

The essential components of a NF installation (Sitaram et al., 2017) include a sensing device (EEG, fNIRS, FMRI, MEG) that captures a signal measuring the RoI activity, a software component analysing NF performance (by comparing the RoI signal to a baseline or reference), and a feedback system which maps the performance measure to a feedback channel giving the user an indication on how well they are activating the target region.

If we leave aside the case of motor areas, most NF experiments require the activation of areas which are not under direct volitional control[2] [for instance, the amydgala (Zotev et al., 2016), insula (Lawrence et al., 2014), PFC (Barth et al., 2016), Anterior Cingulate Cortex (Mathiak et al., 2015; Zilverstand et al., 2017)]. Initially, subjects may use cognitive strategies to facilitate the activation of the target RoI and subsequently guide themselves on the feedback signal to sustain that activation. A cognitive strategy is essentially a set of thought contents which are known to facilitate the activation of the RoI, albeit not always specifically. For instance, imagining a gesture would activate corresponding pre-motor areas, pleasant autobiographic memories can affect PFC asymmetry, concentrating on your inner self may activate the insula (Lawrence et al., 2014), and achieving a relaxed state may decrease the amygdala activity. Cognitive strategies may be suggested by the experimenter, or may be discovered by the subjects themselves based on the (often partial) experimental brief they have been given (Autenrieth et al., 2020). Some of the cognitive strategies may actually lack specificity: for PFC asymmetry in which valence and approach may be confounded, the use of positive autobiographic memories as a cognitive strategy may actually bias activation from a valence perspective. Similarly, different strategies may lead to the same region activation (Lawrence et al., 2014).

Barth et al. (2016) have identified no less than 17 different cognitive strategies used by subjects during a PFC NF experiment (not involving PFC asymmetry). These were often workload-oriented and included verbal fluency tasks, calculating, and naming terms in certain categories. Only two subjects used emotional or arousing strategies, probably due to the fact that there were no affective or motivational element in that PFC experiment: subjects were given visual feedback of their own PFC through an activation heat map, and told to increase activation. Another possible explanation is that subjects may have been influenced in their choice of cognitive strategies by a preliminary working memory task undertaken prior to the actual experiment.

The cognitive strategies adopted by our subjects are specific to each experiment and, short of being dictated by the experimenter, were influenced by the context of the experiment and some high-level instructions given. In the ANG dataset, subjects naturally expressed anger at the designated character, while in the RAP dataset they tried various positive mental attitudes toward the agent. Finally, in the HEU dataset, they had to express motivation or eagerness and developed various strategies, such as mentally encouraging participants in a race.

[2]The following list includes NF experiments based on hemodynamic signals, fMRI or fNIRS.

The NF loop operates by measuring the level of activation, and mapping it onto the feedback signal, so that it reflects in real-time how successful the subject is in activating the RoI. The NF literature, despite its abundance, rarely discusses in-depth this mapping process/function, which is in general a linear mapping between the activation range above the baseline to the variation range of visual feedback. For instance, in our ANG experiment we run preliminary experiments measuring signal variation and define a variation range using the PFC asymmetry signal's standard deviation, in a subject-specific fashion.

There has been growing interest in the nature of the NF feedback channel which can be acoustic (Rosenfeld et al., 1995), or more often, visual. In the latter case, the primary consideration is in the use of abstract symbology or visually realistic signal. To refine this distinction, we propose to categorize the type of visual feedback by taking into account its degree of integration with the interactive application controlled by the BCI (see **Figure 1** for visual details).

Abstract feedback is the dominant modality in NF, and resorts to various gauges or abstract geometrical shapes whose size vary with the activation signal (Trambaiolli et al., 2018), or even screen color (Sakatani et al., 2013). Some feedback can be visually realistic but not semantically related, as in Li et al. (2019). Even feedback based of visualizing target brain areas as in (Barth et al., 2016) should fall in this category. Abstract feedback is primarily used in clinical applications or fundamental NF investigation: in BCI, abstract symbology tends to be disconnected from the main application, unless some metaphor can be established between the abstract shape and an element of the application. For instance, in our HEU dataset, the width of the triangle used for visual feedback is a metaphor of the heuristic search space (Cavazza et al., 2017).

Semantic feedback corresponds to more realistic visual feedback which can relate to the affective signals captured by the BCI signal. For instance, in our ANG dataset visual feedback consists in altering the visibility of the character against which anger has to be expressed after an animation showing his evil nature is shown to the subject.

Finally, **Task-related feedback** refers to experimental conditions in which the BCI input is naturally embedded in the interaction process, for instance with a visual feedback which is part of the interface operation rather than added-on symbology. In our RAP dataset, the overall task consists in non-verbal communication with an agent with the agent's non-verbal behavior actually constituting the feedback signal, making the feedback signal indistinguishable from the task itself. This comes at the cost of losing some real-time properties of the feedback signal, but the benefits of visual realism may actually outweigh this loss by far. In previous work with a different yet compatible signal [fMRI-derived EEG known as electrical fingerprint (Keynan et al., 2019)], we have suggested that some complex visual interfaces may have signal filtering abilities (Yamin et al., 2017).

While visually realistic feedback has been shown to facilitate NF performance via increased competence rather than simple engagement (Cohen et al., 2016) there is no evidence that it would distort the process compared to abstract feedback, which

FIGURE 1 | NF Feedback categorization: abstract [HEU], semantic [ANG], and task-related [RAP].

is why we consider our datasets as representative examples of generic interest. Moreover, visually realistic feedback with a social component (such as in our RAP dataset) has been shown to foster good NF responses even with minimal training (Mathiak et al., 2015). Finally, more speculative explanations could involve improved reward encoding with realistic visual feedback, in some cases even resonating with reward encoding in the RoI itself (Cavazza, 2018), in particular in the case of DLPFC (Tanaka et al., 2006; Aupperle et al., 2015).

NF signals tend to fluctuate significantly during an epoch. There have been several theoretical hypotheses underlying their dynamics, such as the difficulty to activate the region, the difficulty to sustain that activation, and the extent to which feedback could assist or even hinder the process (Hinterberger et al., 2004). Some researchers have hypothesized a control theory model for NF (Ros et al., 2014), in which oscillations would be explained by the response of the controller outside of a steady state mode. BCI uses of NF signal can operate with shorter epochs as no long-term effects are sought, and the actual duration tends to be a compromise between application requirements and signal acquisition.

2. ORIGINS OF DATA

2.1. Common Description of fNIRS NF Experiments

Based on previous literature (Ruocco et al., 2014), including literature applying HbO to affect-related manipulation in the DLPFC (Tuscan et al., 2013), and to approach/withdrawal-related experimental manipulation (Morinaga et al., 2007), and based on our pilot study (Aranyi et al., 2015a), we elected to use HbO for real-time application; we based *post-hoc* analyses on the same metric for consistency.

Note that this measure is relative to a *baseline* (Ayaz et al., 2010), this has important practical consequences in defining and quantifying NF success. For example, as this operationalization of asymmetry yields interval-level data, a ratio of task/no-task signals for defining and quantifying success (for instance Sarkheil et al., 2015) cannot be applied.

In our previous work corresponding to the three datasets ANG, RAP, HEU, we have used a specific terminology in which *baseline* referred to signal value at rest. In some experiments, we used a reference epoch to calculate the signal variation, which in some instances required the subject to watch a similar environment to the one used during NF epochs, sometimes also involving a neutral cognitive task, such as *counting*. It should also be noted that some datasets have considered PFC asymmetry to be zero for the baseline whilst there is evidence of default PFC asymmetry values even in fNIRS (Zohdi et al., 2020) something which was readily captured in EEG experiments (Cavazza et al., 2014) but needs to be redefined on a session or even epoch basis when using hemodynamic signals.

In all these experiments NF is used for its ability to produce a signal with a clear interpretation in cognitive terms. We are using PFC asymmetry as a dimensional marker of approach, the actual cognitive feature under consideration (or analysis) being determined by the experimental context, and the nature of the feedback signal. For instance, an experiment on anger will measure approach (dissociated from valence) (Harmon-Jones, 2007), while in an experiment on empathy, approach can be used as a proxy measurement (Cavazza et al., 2014).

It is worth discussing again the main differences between clinical NF and BCI NF, the latter still being an emerging application within the broader field of BCI technology. In terms of experimental protocol and validation, clinical NF tends to rely on sham feedback as a control group, under the hypothesis that appropriate feedback provides the reward signal that mediates long-term effects. The clinical context implies and allows the use of repeated sessions with significant training, which increases the number of responders: on the other hand, BCI NF dedicates limited time to subjects training and leaves non-responders to the various categories of BCI illiteracy (Lee et al., 2014; Trambaiolli et al., 2018; Autenrieth et al., 2020), concentrating instead on the responders' behavior.

2.2. Representativity and Interest of the Dataset

In this section, we are briefly discussing the relevance and potential community interest of our datasets, considering the increasing popularity of fNIRS and in particular fNIRS NF. We will be basing this discussion primarily on the recent review of fNIRS NF by Kohl et al. (2020), which collected a significant

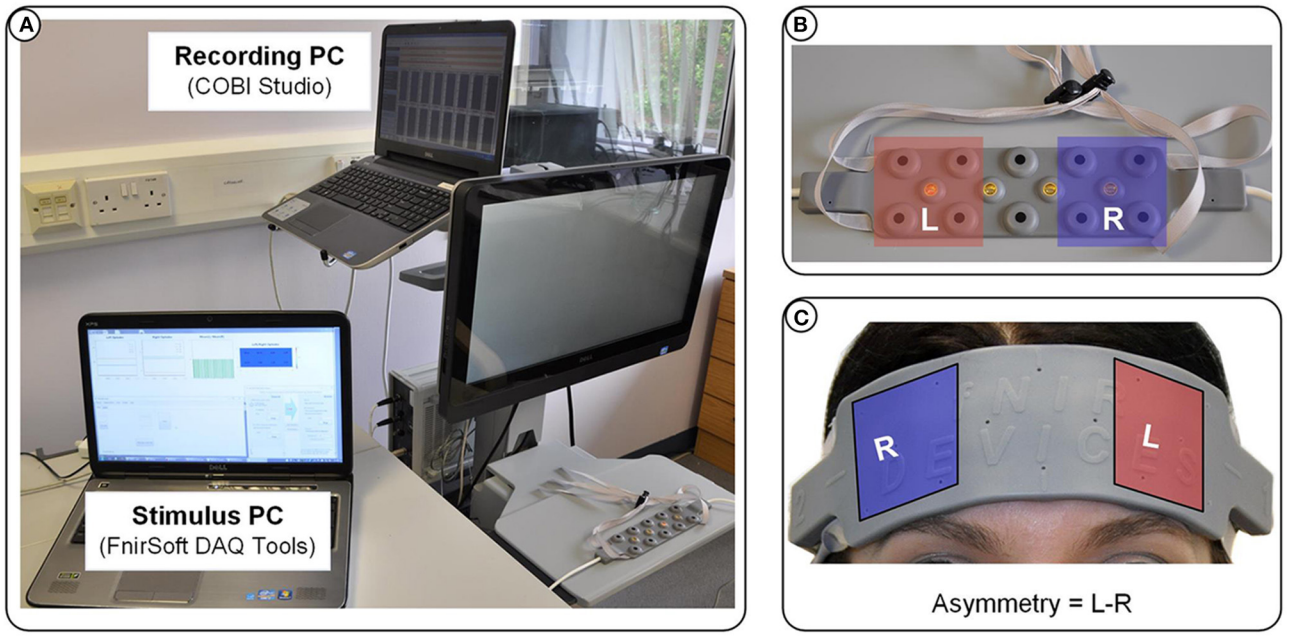


FIGURE 2 | (A) Overview of the apparatus used in our experiments: fNIR Optical Brain Imaging System (fNIR400) by Biopac Systems with one PC dedicated to the data acquisition, and one PC dedicated to running the simulation and visualization of the stimulus presented in real-time to the subjects. **(B)** The 16-channel sensor placed on the subjects' forehead **(C)** showing the selected channels for the calculation of the asymmetry values.

number of studies and highlights variants in different core aspects of NF[3].

Firstly, most of the studies reported in the review use HbO, which is also the case for our three datasets (Aranyi et al., 2015a). Recently (Tachtsidis and Scholkmann, 2016) have suggested that HbO alone might be insufficient to cover the widest range of experimental situation, but this recent observation has not yet been fully taken up in the community.

The duration of NF epoch in Kohl et al. (2020) ranges from 5 to 40 s (the latter actually corresponding to our own RAP dataset, although its actual useful duration is 33 s), with the majority of epochs (30%) lasting 30 s. Duration of epochs in our datasets are 30 s [HEU], 15 s [ANG], and 33 s [RAP].

Our subject population was primarily drawn from healthy subjects, which were also overrepresented in Kohl et al. (2020)'s review (76%). Our target RoI also proves to be one of the most studied ones, as 59% of the studies reviewed trained participants to regulate parts of the PFC.

Our three datasets also cover a range of cognitive strategies, ranging from explicitly expressing a given feeling [ANG], engaging with a virtual character [RAP], or expressing motivation [HEU]. We collected post-experiment user feedback (Autenrieth et al., 2020) on the actual cognitive strategies they used for NF: it highlighted a mix of through contents related to both approach and positive valence, apparently influenced by the visual nature of the application. Users reported various sorts of mental "cheering" as if encouraging runners during a race, as well as the use of more abstract thinking strategies to generate a feeling of eagerness, such as reminiscence of appetitive stimuli or pleasant memories.

[3]This review already includes one of our own experiments corresponding to the RAP dataset.

3. DETAILED EXPERIMENTS

3.1. Apparatus

For the three experiments presented here, we used an fNIR Optical Brain Imaging System (fNIR400) by Biopac Systems for data acquisition. Raw fNIRS data and oxygenation values were collected with 2 Hz sampling rate (using COBI Studio and FnirSoft), and was sent to the bespoke experimental software over TCP/IP (using FnirSoft DAQ Tools) (**Figure 2A**). A 16-channel sensor with a fixed 2.5 cm source-detector separation was placed on the subjects' forehead. For real-time application we used measurements of changes in HbO concentration, as opposed to deoxygenated or total hemoglobin (HbR and HbT, respectively) (**Figures 2B,C**). Values of HbO concentration changes were averaged over the four leftmost and four rightmost channels (located over the left and right DLPFC, respectively) to derive a simple metric of inter-hemispheric difference in the level of HbO change that could account for left prefrontal asymmetry (i.e., $Asymmetry = L - R$).

A bespoke graphical user interface was developed using C# and Windows WPF for each of the experiment, which includes a real-time visualization of HbO changes and asymmetry values. All these variables are logged during the experiment to facilitate the post-processing of the collected data. This graphical user interface was also used to manage the running of the experiment

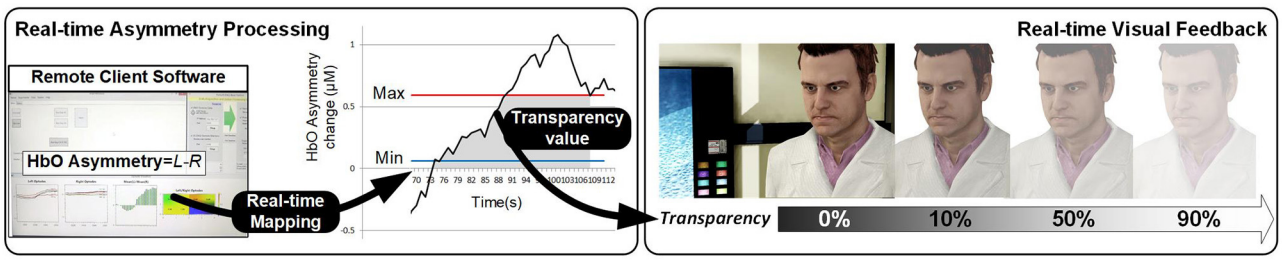


FIGURE 3 | Experimental setup for the ANG experiment (see **Figure 1** [ANG] for overall visual feedback setup).

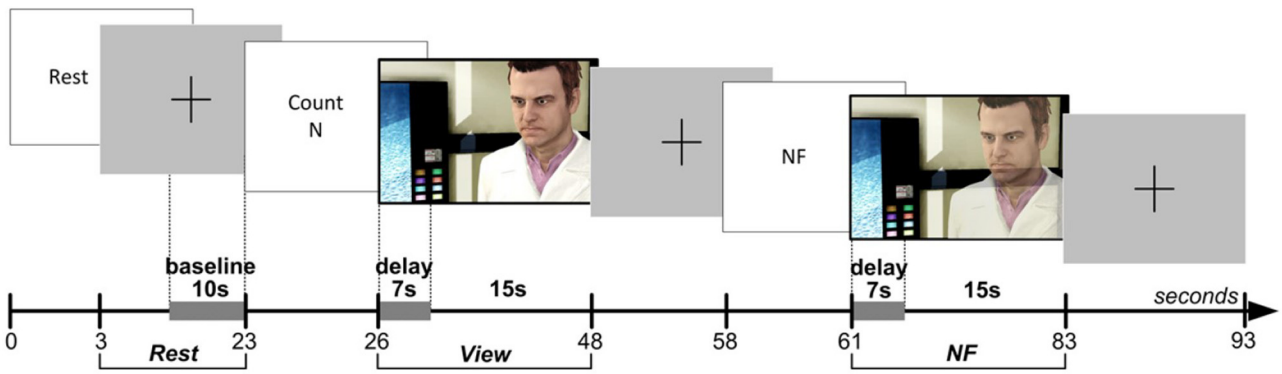


FIGURE 4 | Protocol design for the ANG experiment.

itself, which included displaying the required epochs (text, image, video, or more complex visualization) and managing their duration and the synchronization of all the software components. Finally, this interface was also used to implement visual feedback.

For the positioning of the different devices in relation to the subjects, we followed the recommendations of Solovey et al. (2009) regarding the use of fNIRS in a HCI setting. Subjects were seated ∼47″ (120 cm) from a 24″ flat monitor in a dimly-lit, quiet (but not soundproof) room in a comfortable chair to minimize movements, with the fNIRS probe positioned over their forehead and covered with non-transparent fabric to prevent ambient light reaching the sensors. Subjects were instructed to refrain from moving their limbs, frowning and talking during data collection blocks.

3.2. Anger-Based NF—[ANG]

3.2.1. Subjects

This experiment (Aranyi et al., 2015b) was conducted with twelve English-speaking adult subjects originally, though one subject had to be excluded due to technical problems. Thus, the effective sample size was eleven subjects (five females, mean age = 33.55 years, SD = 11.53, range: [24; 59]). Subjects had no history of psychiatric conditions and were right-handed. They all provided written consent prior to participation.

3.2.2. Protocol

Full details of the protocol design can be found in Aranyi et al. (2015b), but we will outline the essential details here (**Figure 3** for the details of the overall setup). Subjects were instructed that they would go through a sequence of blocks, each comprising three main epochs: Rest, View and NF (**Figure 4**). During Rest, the baseline for calculating HbO data is acquired for the block as a whole. The View epochs correspond to control conditions, during which the subjects watch an idle animation of the character while given a cognitive task, *counting*, that keeps them in a neutral state, hence providing a reference for prefrontal asymmetry levels. Finally, the NF epochs consist in subjects expressing anger toward the character and receiving visual feedback. Each subject completed two training blocks to get acquainted to the task, followed by six blocks for the experiment itself, during which HbO asymmetry was monitored and recorded.

3.2.3. Results

We treated a block as successful if the mean of asymmetry values during the NF epoch was statistically significantly larger than the mean of asymmetry values during the View epoch within the same block (**Figure 5**). As opposed to simply comparing asymmetry scores during NF to the overall baseline, we compared asymmetry scores between the successive View and NF epochs because the visual stimulus was very similar (and conceptually the same) in the two epochs. Moreover, we set conditions for controlling subjects' cognitive activity during these epochs

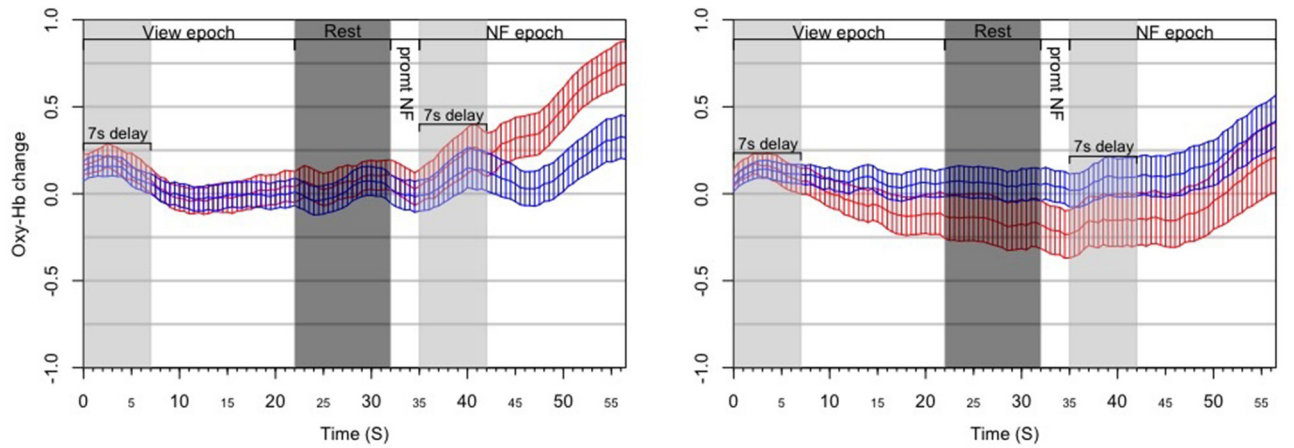


FIGURE 5 | Results from the *post-hoc* analysis of the ANG experiment, illustrating the dynamics of PFC asymmetry over the whole experiment for *successful* blocks **(Left)** and *unsuccessful* blocks **(Right)** successful blocks demonstrates a significant increase of the left-side oxygenation compared to unsuccessful blocks.

(counting during View and expressing anger during NF), whereas thought processes during Rest were not controlled. Thus, the View epoch served as a control condition within each block. Because the hemodynamic response measured by fNIRS occurs in ∼7 s, we discarded the first 7 s of data in each View and NF epoch for determining block success. The system determined block success by performing an independent *t*-test on the set of asymmetry scores collected during successive View and NF epochs within a block. In particular, it calculated mean and standard deviation of asymmetry scores in both epochs, and then calculated the t value. Since removing the first 7 s left 15 s of data per epoch (at least 29 data points sampled at 2 Hz), the software used the t critical value of 2.05 with 28 degrees of freedom for p (two-tailed) = 0.05 as a threshold for success. Furthermore, to quantify the extent of block success by expressing the distance of the distribution of asymmetry scores during successive View and NF epochs, the experimental software calculated the Cohen's d effect-size measure, which is the difference between two means divided by the pooled standard deviation. This way we characterized each block with a dichotomous success value (success/fail) and a continuous success score (Cohen's d or d for short) that reflects the distance between the distribution of asymmetry values between View and NF epochs within the same block.

3.3. Virtual Agent—[RAP]

3.3.1. Subjects

This experiment (Aranyi et al., 2016) was conducted with eighteen English-speaking adult subjects, though data from one subject was discarded due to technical problems during data collection. Thus, the effective sample size was seventeen subjects (eight females, mean age = 35.11 years, SD = 11.25, range: [21; 60]). Subjects were right-handed and were not treated for psychiatric conditions.

3.3.2. Protocol

Figure 6 provides an overview of the details of the overall setup. The experimental task consisted in completing eight identical blocks (preceded by a practice block which was not analyzed). The structure of the blocks is presented in **Figure 7**. Each block included three epochs: Rest, View, and Engage. During Rest epochs, subjects were instructed to look at a crosshair located in the center of a gray screen to try to clear their head of thoughts and relax. During View epochs, subjects were instructed to keep looking at the agent while carrying out a simple mental counting task (counting backwards from 500 by increments of a given integer). This task was included to control for unwanted mental processes. During Engage epochs, subjects were instructed to engage with the ECA through positive thinking, and to "cheer her up" with their thoughts. We were deliberately vague with support instructions in order to allow subjects to develop their own cognitive strategies. After completing each block, subjects were asked to describe their strategies in general terms. During Engage epochs, subjects received real-time feedback of their left-asymmetry. To ensure consistent mapping of individual variations in left-asymmetry onto the feedback signal, we used the range of variation of HbO asymmetry during the View epoch in each block to determine the mapping of the level of engagement from the user to the visual feedback signal. This was calculated by the experimental software during the last 3 s of the Rest epoch between the View and Engage epochs. We defined the minimum point for mapping Min as the mean of left-asymmetry values during the View epoch plus 1.28 times their standard deviation. In normally distributed asymmetry scores, this threshold would result in no feedback for 90% of the spontaneous asymmetry variations during the reference (View) epoch. To determine the maximum Max point for mapping, we increased the threshold asymmetry value for feedback Min by the variation range of asymmetry values during the View epoch. Asymmetry values within the range [Min; Max] during the Engage epoch were mapped linearly onto the ECA's facial expression, with the same 2Hz frequency as the acquisition of asymmetry values.

3.3.3. Results

Subjects were instructed to refrain from talking, frowning and moving their limbs during fNIRS data collection periods within

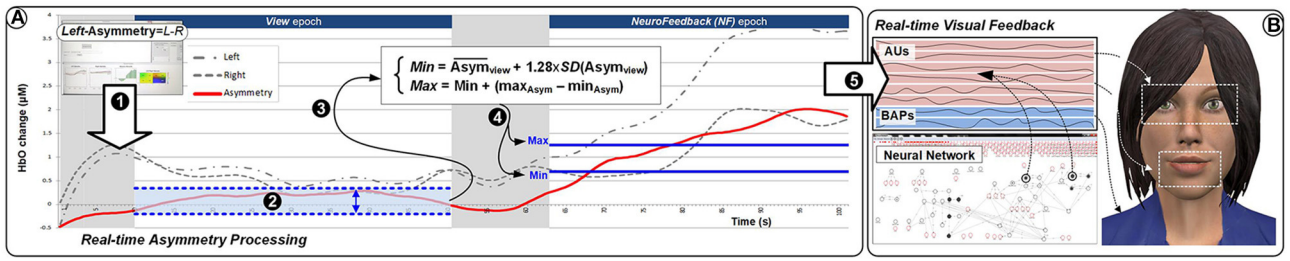


FIGURE 6 | (A,B) Experimental setup for the RAP experiment (see **Figure 1** [RAP] for overall visual feedback setup).

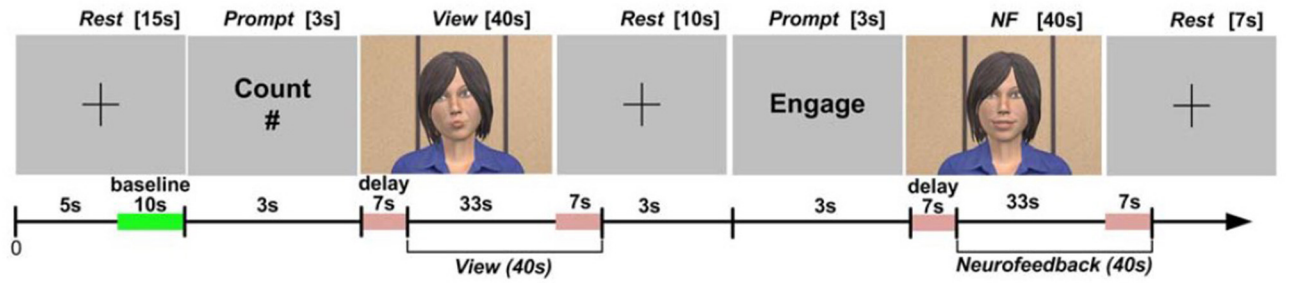


FIGURE 7 | Protocol design for the RAP experiment.

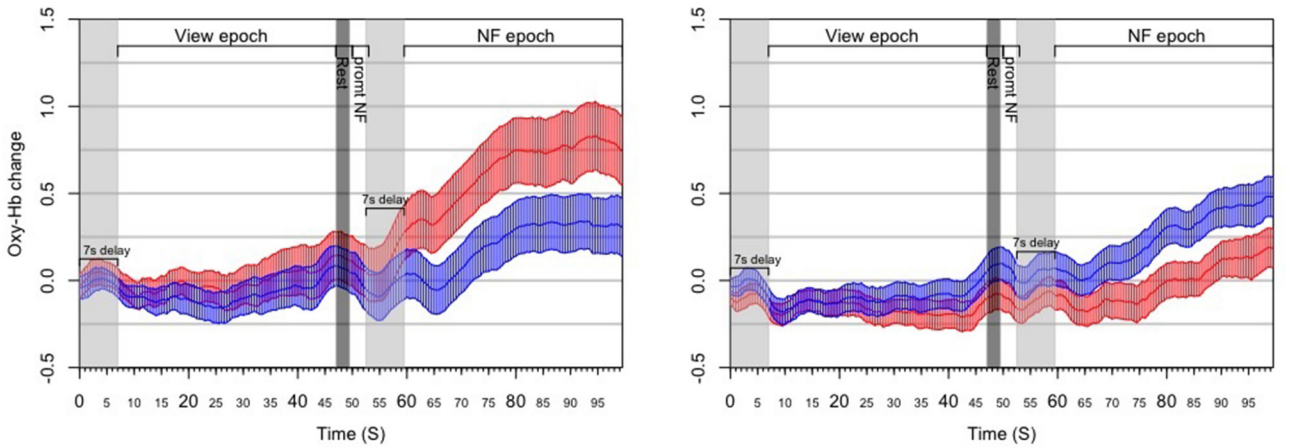


FIGURE 8 | Results from the *post-hoc* analysis of the RAP experiment, illustrating the dynamics of PFC asymmetry over the whole experiment for *successful* blocks **(Left)** and *unsuccessful* blocks **(Right)** successful blocks demonstrates a significant increase of the left-side oxygenation compared to unsuccessful blocks.

the protocol. Additionally, we applied a sliding-window motion artifact rejection (SMAR) to each channel used for calculating the asymmetry metric which was inspected *post-hoc* to identify motion artifacts during NF. For *post-hoc* analyses, raw data were low-pass filtered using a finite impulse response filter with order 20 and 0.1 Hz cut-off frequency (Ayaz et al., 2010). For this experiment (Aranyi et al., 2016), we used a sliding-window motion artifact rejection (SMAR) procedure, which rejected motion-affected periods in the fNIRS signal. This was an experiment which had greater potential for upper body movement as the subject could try to align to the agent non-verbal behavior which also included head and upper body motion.

Moreover, we have applied a counting task during reference epochs rather than passive visualization: while counting tasks are known to activate the PFC (Barth et al., 2016) they are also neutral toward affective and motivational aspects, which allows us to claim greater specificity for measuring NF activation compared to our reference epoch (**Figure 8**).

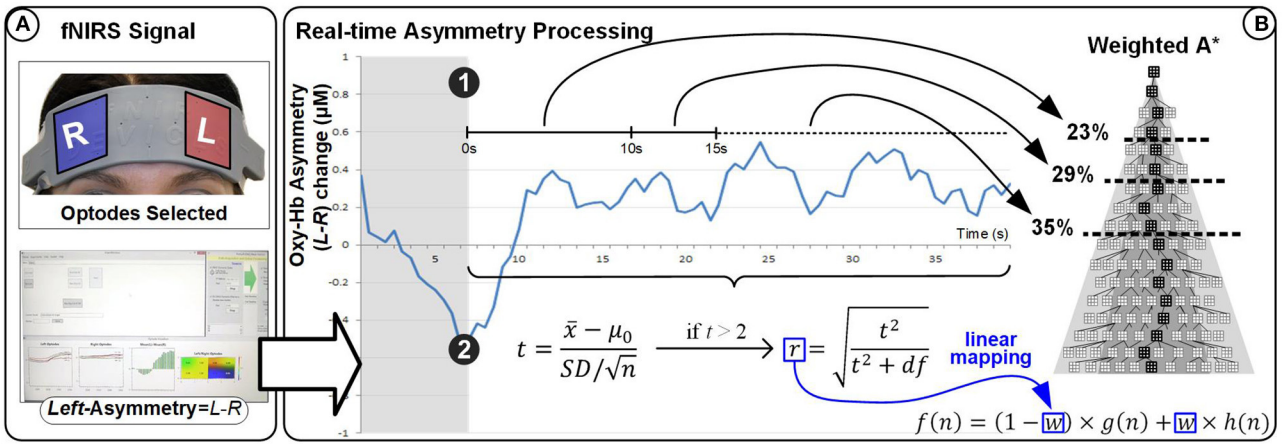


FIGURE 9 | Experimental setup for the HEU experiment (see **Figure 1** [HEU] for overall visual feedback setup). A* represent is artificial intelligence search algorithm.

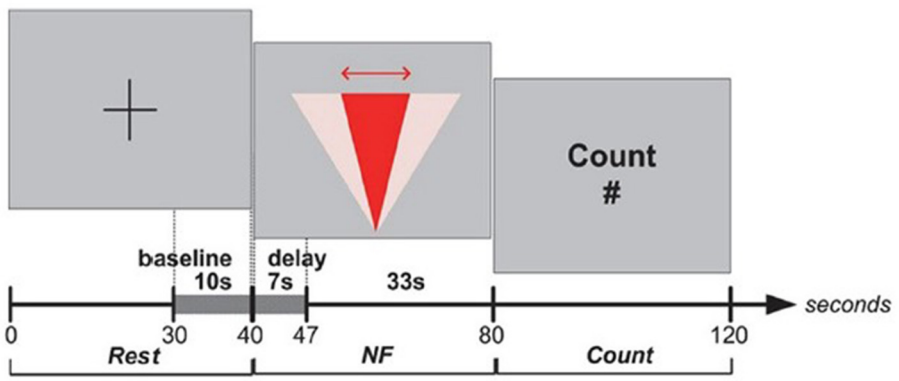


FIGURE 10 | Protocol design for the HEU experiment.

3.4. Motivational BCI—[HEU]

3.4.1. Subjects

This experiment (Cavazza et al., 2017) was conducted with eleven adults (three females; mean age = 37.18 years, SD = 11.21, range = [20; 52]) who were right-handed, reported no treatment history for psychiatric conditions and provided written consent prior to participation. Subjects were seated in a dimly-lit room in a comfortable chair to minimize movements, with the fNIRS probe positioned over their forehead and covered with non-transparent fabric to prevent ambient light reaching the sensors.

3.4.2. Protocol

HbO values were averaged over the four leftmost and the four rightmost channels (located over the left and right dorsolateral prefrontal cortex, respectively). Average right HbO was subtracted from average left HbO to derive a simple, real-time prefrontal asymmetry score rejecting differential changes in oxygenation. We developed bespoke experimental software for generating real-time feedback and interfacing with the WA* algorithm (**Figure 9**). Response time is an important component of NF systems; however, (Zotev et al., 2016) reported successful fMRI-based NF despite the ∼7 s delay of the BOLD signal. Since delay using fNIRS is comparable, we sought inspiration from the experimental protocol of (Zotev et al., 2016). The overall protocol design for the experiment is described in Cavazza et al. (2017) and in **Figure 10**.

3.4.3. Results

Out of all 66 blocks completed by the eleven subjects, 38 (58%) contained an NF epoch with statistically significant left-side asymmetry; these blocks were considered successful. Each subject had at least one successful block, and eight subjects (73%) had at least three successful blocks (i.e., half of blocks successful). No subject achieved NF success on all six blocks. Since fNIRS signals are relative values, it can be difficult to compare them across subjects (Sakatani et al., 2013); moreover, the magnitude of oxygenation changes can also differ substantially across blocks within the same subject (**Figure 11**). Our mapping strategy was designed to mitigate the issue of comparability.

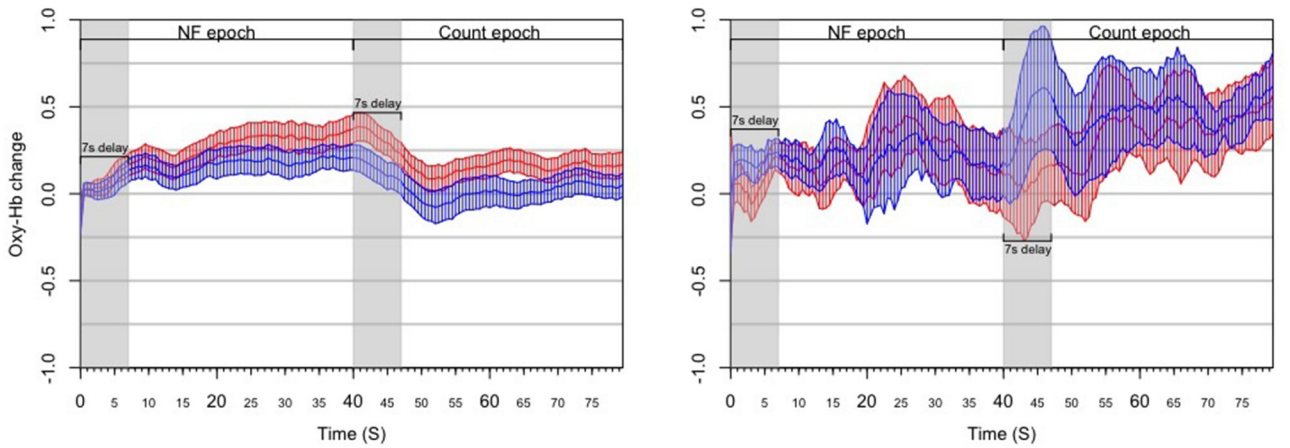


FIGURE 11 | Results from the *post-hoc* analysis of the HEU experiment, illustrating the dynamics of PFC asymmetry over the whole experiment for *successful* blocks **(Left)** and *unsuccessful* blocks **(Right)** successful blocks demonstrates a significant increase of the left-side oxygenation compared to unsuccessful blocks.

4. DATA FORMATS AND DATASETS

4.1. Data Formats

The benefits of offering datasets to the BCI community must allow for the data to be easily manageable by all, which includes requirements, such as the ability to process the data with a wide range of modern software, in our case this includes Matlab or R, and also the ability to account for latest offerings in terms of programming languages, such as Python or Julia.

Other initiatives have supported the exchange of NIRS data, for instance the Shared Near Infrared Spectroscopy Format (SNIRF)[4], developed by the Society for functional Near Infrared Spectroscopy. In an effort to facilitate the sharing of NIRS data, they have developed the Shared Near Infrared Spectroscopy Format (SNIRF). It follows a hierarchical data format—HDF5 which is a general purpose, machine-independent standard for storing scientific data in files, developed by the National Center for Supercomputing Applications (NCSA). As well as SNIRF, they have also developed two other text-based alternatives for platforms that do not support HDF5—JNIRS and BNIRS which are JSON and binary JSON files with the forementioned file extensions.

We have opted for a more portable and lightweight alternative to HDF5: JSON. To facilitate access to our data, we also provide binary data files for the most commonly used scientific languages—R, MATLAB, Python and Julia. The binary data allows for instant access to the data without the prerequisite of being familiar with HDF5 or JSON and acquainted with the necessary libraries that are needed to load the data in any given environment.

[4]https://github.com/fangq/snirf/blob/master/snirf_specification.md

Unlike SNIRF, which follows a generic filing-like hierarchical data structure, we have taken an object-oriented approach to structuring our data. This approach is self-contained and is descriptive of our experiments as we have objects that define a subject, an experimental block and an epoch (**Figure 12**). This approach facilitates the understanding and possible analysis of the data, as we have included as many properties as needed for ease of use and to accurately depict our experiments, such as a Boolean value property which indicates whether an experimental block was characterized as successful or not, the unfiltered HbO values for each channel used and the filtered asymmetry scores used during the NF epoch—and many more.

Our intention was to provide datasets ready for use, i.e., requiring minimum data wrangling prior to analysis. Though SNIRF provides a generic mechanism to share NIRS data, one would still need to extract relevant information and restructure the data depending on the intended data analysis. Our data format also reflects the fact that we are sharing fNIRS data in a NF context, and supports additional annotations typical of NF on top of RoI signal dynamics (in this case, PFC activity).

Thus, we decided to provide our datasets in multiple file formats so as to expand the usability of our datasets across software and languages—we therefore provide our data as files for MatLab (file extension *.mat*), R (file extension *.RData*), JSON (file extension *.json*), Pickle (file extension *.pckl*), and Julia (file extension *.jld*).

In this section, we provide details of the overall structure of the datasets, as well as highlighting detailing specificities, in order to make them as accessible as possible for further processing and analysis using programming languages Python or Julia, or from recognized data processing and analysis packages, such as Matlab and R. We first present generic data which are common to all our experiments, then we provide details about the specific Subjects data, as well as Blocks and Epochs, and finishing with a short discussion on Time Series (see framework and structural datasets details in **Figure 12**, as well as a summary of both the experiments

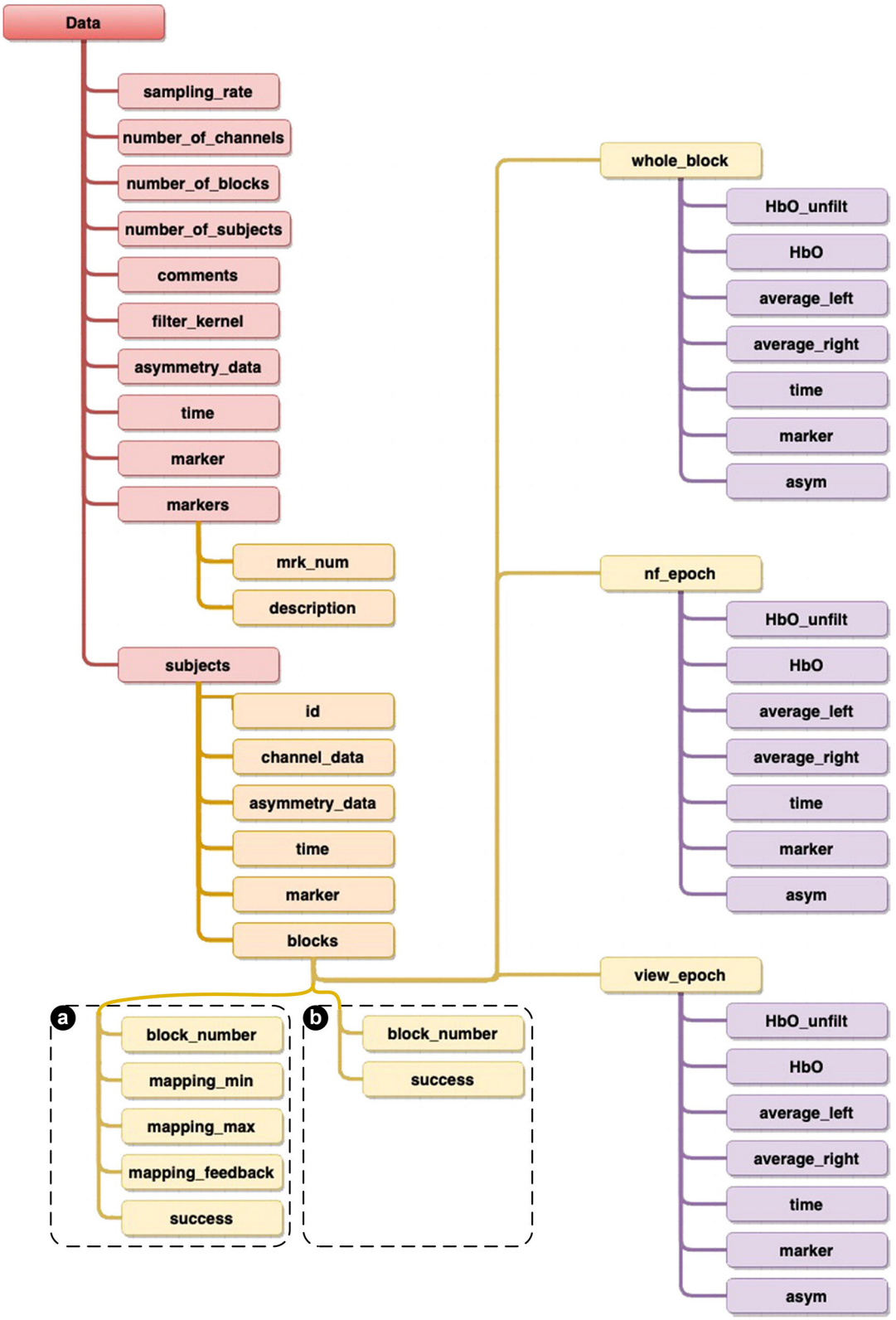


FIGURE 12 | Diagram providing details of the structure of the overall datasets. The only difference in structures are shown in the blocks data, where **(a)** is specific to both [ANG] and [RAP] experiments, whilst **(b)** is specific to the [HEU] experiment.

TABLE 1 | Comparison table of overall protocol and analysis settings for the three experiments.

Experiment	HEU	ANG	RAP
Threshold	0	Dynamic (M + 1.28*SD)	Dynamic (M + 1.28*SD)
Maximum	1.1 (fixed)	Dynamic (min + range)	Dynamic (min + range)
Practice	3 blocks	2 blocks	1 block
N blocks	6	6	8
Baseline task	Rest	Rest	Rest
Reference epoch	No	Yes	Yes
Test	Parametric	Parametric	Bootstrapping
Success test	Real-time	Real-time	*Post-hoc*
Filtering	No	No	Yes (FIR, SMAR, detrending)
Success measure	r	d	r
Delay treatment	Remove 7 s	Remove 7 s	Windowing
Effective NF epoch length	33 s	15 s (+15 s reference task)	40 s (+40 s reference task)
N subjects	11	11	17
Block success	58%	58%	58%
Subject success	73%	73%	70%

TABLE 2 | Comparison table of participants' demographic information for the three experiments.

Experiment	HEU	ANG	RAP
N subjects	11	11	17
N female subjects	3	5	8
Mean age	37.18	33.55	35.11
SD age	11.21	11.53	11.25
Range	[20; 52]	[24; 59]	[21; 60]

settings in **Table 1** and participants' demographic information in **Table 2**).

4.2. Generic Data

4.2.1. Sampling Rate and Number of Channels

By default, our fNIR system records two wavelengths and dark current for each 16 optodes, totalling 48 measurements for each sampling period. Although the sampling rate of the latest generation systems can be up to 10 Hz, our experiments were recorded at 2Hz sampling rate (*Sampling_rate*). Although the fNIR system provides full data for all 16 optodes, as we only consider data for the calculation of the asymmetry scores, we here provide HbO data for eight optodes (*Number_of_channels*). Asymmetry calculations are derived from the four leftmost and four rightmost optodes, as a difference of Left minus Right (see **Figures 2B,C**).

4.2.2. Asymmetry Data and Filtering

The asymmetry data (*Asymmetry_data*) calculated in real-time during the experiments is provided in the datasets as a tensor in the following format (subject, time, blocks). This subset constitutes the asymmetry data for all subjects and all blocks in the experiment considered. Each epoch within has been resampled and the data has been filtered using FIR filter kernel (*Filter_kernel*) applied as a low-pass filter to each raw channel data.

4.2.3. Time and Blocks

All data provided includes timestamps in the form of a normalized time vector (*Time*) which is the temporal reference for all data in the experimental blocks considered. As described in detail for each protocol design in the previous section, we also record the number of experimental blocks for each subject (*Number_of_blocks*). These individual subject blocks also include the practice blocks. Practice blocks were actual blocks, as per the experimental design, used as a prior task for the subjects to acquaint themselves with the task which was expected from them. Subjects were not directly influenced by the experimenter during these blocks, but were given the opportunity to discover what the realtime system consisted of, overall. Although, practice blocks were also logged in our original data, we do not include them here, since they were not part of our original analysis.

4.2.4. Markers

All our experiments include the very important markers data (*Markers*), which has been set consistently across our three experiments. It is provided as a collection of marker IDs, which include the "num," being the marker number used to label the time series, as well as a "description" providing details of the epoch they refer to (e.g., 54 is "NF epoch"). These marker IDs are generated during the experiments so as to facilitate the extraction of the blocks from the raw data for the post-analysis. Marker data (*Marker*) is provided as a vector of the markers for each data point, so as to be able to label individual data points with the allocated reference marker, which then provides the details of which epoch they refer to.

4.3. Subjects

The all-important data recorded per subject is being made available on a per subject basis. For each subject we provide a reference (*id*) which is the subject's unique identifier. This identifier was generated following the convention "MDDN," where M is the number of the month of the day the experiment took place, DD is the day of the experiment, and N is the subject

order on the day of the experiment—(i.e., 3,241 is the 1st subject to take part in the experiment on the 24th of March. This method of anonymizing the subjects' details were deemed sufficient to be able to retrieve the subject's data if they had decided, at any point, to retract their data from the experiment analysis. (*Channel_data*) is a tensor in the format (channel, time, blocks) containing the channels for all blocks. Each epoch within has been resampled and the data has been filtered. The asymmetry information provided (*Asymmetry_data*) is a matrix in the format (time x blocks) containing the asymmetry data for all blocks. Each epoch within has been resampled and the data has been filtered. As expected from any data information, time stamps are also provided for each data entry (*Time*), which is a normalized time vector providing temporal labeling for the "Channel data" and "Asymmetry data" time series. Then, as presented previously, we also include marker data (*Marker*) as a vector of markers for segmenting the "Channel data" and "Asymmetry data" time series. And finally, (*Blocks*) is a collection of all blocks data. This is presented in the next section.

4.4. Blocks

Blocks are characterized by their identifier (*Block_number*) which describes the order of the block (starting at zero). We define a boolean variable (*Success*) indicating whether the block was successful based on the actual success criteria defined for the considered experiment (refer to the above sections for details of the definition of the success criteria). (*Whole_block*) is a variable which contains the data for the whole block (details are provided in the next section on Epochs). As we have presented in the details of the experiments above, the two important epochs are provided here as View (*View_epoch*), being a variable which provides the details of the View epoch data only, after having been segmented from the *Whole_block*, and the NF epoch data (*NF_epoch*), being a variable which provides the details of the NF epoch data only, after having been segmented from the *Whole_block*. Contrary to the HEU experiment, in both the ANG and RAP experiments, mapping is defined between the realtime asymmetry value processed and the feedback value, which is calculated on the basis of a [Min; Max] range—determined in realtime from the reference View epoch. Thus these two experiment datasets also include the following information for each block:

- (*Mapping_min*) is the value calculated as the lower bound value as: the Mean of the View epoch + 1.28 standard deviation of View epoch.
- (*Mapping_max*) is the value calculated as the upper bound value as: the Mapping min value + range of the View epoch.
- (*Mapping_feedback*) is the actual NF signal value mapped to the feedback.

4.5. Epochs

After having described the details of the actual structure of the data presented in the blocks, we are providing details for the epoch data points. We note that *Whole_block*, *View_epoch*, and *NF_epoch* follow the same data structure. (*Time*) is the unnormalized time vector containing the exact time each data point was recorded during the experiment. This provides accurate and detailed overview of the actual recordings and affords possibilities for the potential further analysis. (*Marker*) is a vector of markers allowing for the segmentation of time, HbO filtered and unfiltered, Average left, Average right, and asymmetry values. (*HbO*) provides a matrix of the low-pass filtered HbO channel data in the format channel x time. (*HbO_unfiltered*) provides the same matrix format for the unfiltered HbO channel data. (*Average_left*) is the average value of the four leftmost optodes. (*Average_right*) is the average value of the four rightmost optodes. And finally, (*Asymmetry*) is the actual asymmetry value generated as the difference of *Average_left* and *Average_right*.

4.6. Time Series Discussion

The ANG dataset differentiates between approach and valence, albeit not perfectly, and could be used to experiment whether differences of magnitude take place by removing the valence component. In the HEU dataset, NF success above baseline is used primarily as a trigger so could be of interest on comparative study of NF dynamics but perhaps less on NF epoch-based validation. The RAP dataset is closer to previous EEG experiments (Cavazza et al., 2014) and the one with perhaps the most potential for confounding various aspects of PFC asymmetry in terms of its dimensional interpretation (approach, valence). On the other hand, it has some of the longest fNIRS NF epochs (Kohl et al., 2020) and is a good candidate to study signal dynamics.

5. CONCLUSIONS

We have described three datasets for fNIRS PFC asymmetry, which correspond to one of the most investigated signals in social and affective neuroscience and also one of the main areas for fNIRS NF. Although the focus of our BCI experiments were primarily on the motivational dimension, the DLPFC signals can also be of interest to researchers requiring comparative data when investigating cognitive workload or other dimensions, such as valence. As these datasets cover different NF variants, they should be valuable to investigate signal dynamics across epochs of different lengths as well as issues around baselining and reference epochs. Since they all have been previously analyzed as part of various publications (Aranyi et al., 2015b, 2016; Cavazza et al., 2017), they can also support experiments with various statistical methods for *post-hoc* epoch validation. We have endeavored to facilitate this through the various formats we have embedded data into, which should support various processing pipelines in data analysis or machine learning.

DATA AVAILABILITY STATEMENT

The datasets generated for these studies can be found in the BCI NeuroFeedback fNIRS repository: https://github.com/fcharles-BU/BCI-NF-fNIRS.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Teesside University Research Ethics Committee, Middlesbrough, Tees Valley, TS1 3BX, UK. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

FC contributed to the data collection, systems development, and writing of the manuscript. CD contributed to the dataset generation and analysis, and writing of the manuscript. MC contributed to the original experimental designs and writing of the manuscript. All authors have made direct and substantial intellectual contributions to the article and approved it for publication.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Gabor Aranyi for his essential contribution to the proof-of-concept experiments and data collection, whose results are presented in the experiments discussed in this paper.

REFERENCES

Afergan, D., Shibata, T., Hincks, S. W., Peck, E. M., Yuksel, B. F., Chang, R., et al. (2014). "Brain-based target expansion," in *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, HI), 583–593. doi: 10.1145/2642918.2647414

Aranyi, G., Cavazza, M., and Charles, F. (2015a). "Using fNIRS for prefrontal-asymmetry neurofeedback: methods and challenges," in *International Workshop on Symbiotic Interaction* (Cham: Springer), 7–20. doi: 10.1007/978-3-319-24917-9_2

Aranyi, G., Charles, F., and Cavazza, M. (2015b). "Anger-based BCI using fNIRS neurofeedback," in *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (Charlotte, NC), 511–521. doi: 10.1145/2807442.2807447

Aranyi, G., Pecune, F., Charles, F., Pelachaud, C., and Cavazza, M. (2016). Affective interaction with a virtual character through an fNIRS brain-computer interface. *Front. Comput. Neurosci.* 10:70. doi: 10.3389/fncom.2016.00070

Aupperle, R. L., Melrose, A. J., Francisco, A., Paulus, M. P., and Stein, M. B. (2015). Neural substrates of approach-avoidance conflict decision-making. *Hum. Brain Mapp.* 36, 449–462. doi: 10.1002/hbm.22639

Autenrieth, M., Kober, S. E., Neuper, C., and Wood, G. (2020). How much do strategy reports tell about the outcomes of neurofeedback training? A study on the voluntary up-regulation of the sensorimotor rhythm. *Front. Hum. Neurosci.* 14:218. doi: 10.3389/fnhum.2020.00218

Ayaz, H., Izzetoglu, M., Shewokis, P. A., and Onaral, B. (2010). "Sliding-window motion artifact rejection for functional near-infrared spectroscopy,"in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology* (Buenos Aires: IEEE), 6567–6570. doi: 10.1109/IEMBS.2010.5627113

Bak, S., Park, J., Shin, J., and Jeong, J. (2019). Open-access fNIRS dataset for classification of unilateral finger-and foot-tapping. *Electronics* 8:1486. doi: 10.3390/electronics8121486

Barth, B., Strehl, U., Fallgatter, A. J., and Ehlis, A.-C. (2016). Near-infrared spectroscopy based neurofeedback of prefrontal cortex activity: a proof-of-concept study. *Front. Hum. Neurosci.* 10:633. doi: 10.3389/fnhum.2016.00633

Cartocci, G., Maglione, A. G., Modica, E., Rossi, D., Canettieri, P., Combi, M., et al. (2016). "The "neurodante project": neurometric measurements of participant's reaction to literary auditory stimuli from dante's "divina commedia"," in *International Workshop on Symbiotic Interaction* (Cham: Springer), 52–64. doi: 10.1007/978-3-319-57753-1_5

Cavazza, M. (2018). A motivational model of BCI-controlled heuristic search. *Brain Sci.* 8:166. doi: 10.3390/brainsci8090166

Cavazza, M., Aranyi, G., and Charles, F. (2017). BCI control of heuristic search algorithms. *Front. Neuroinform.* 11:6. doi: 10.3389/fninf.2017.00006

Cavazza, M., Aranyi, G., Charles, F., Porteous, J., Gilroy, S., Klovatch, I., et al. (2014). "Towards empathic neurofeedback for interactive storytelling," in *2014 Workshop on Computational Models of Narrative* (Dagstuhl: Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik). doi: 10.4230/OASIcs.CMN.2014.42

Coan, J. A., and Allen, J. J. (2002). "The state and trait nature of frontal EEG asymmetry in emotion," in *The Asymmetrical Brain* (MIT Press). doi: 10.7551/mitpress/1463.003.0023

Cohen, A., Keynan, J. N., Jackont, G., Green, N., Rashap, I., Shani, O., et al. (2016). Multi-modal virtual scenario enhances neurofeedback learning. *Front. Robot. AI* 3:52. doi: 10.3389/frobt.2016.00052

Davidson, R. J. (2004). What does the prefrontal cortex "do" in affect: perspectives on frontal eeg asymmetry research. *Biol. Psychol.* 67, 219–234. doi: 10.1016/j.biopsycho.2004.03.008

Ernst, L. H., Plichta, M. M., Lutz, E., Zesewitz, A. K., Tupak, S. V., Dresler, T., et al. (2013). Prefrontal activation patterns of automatic and regulated approach-avoidance reactions-A functional near-infrared spectroscopy (fNIRS) study. *Cortex* 49, 131–142. doi: 10.1016/j.cortex.2011.09.013

Fishburn, F. A., Norr, M. E., Medvedev, A. V., and Vaidya, C. J. (2014). Sensitivity of fNIRS to cognitive state and load. *Front. Hum. Neurosci.* 8:76. doi: 10.3389/fnhum.2014.00076

Harmon-Jones, E. (2007). Trait anger predicts relative left frontal cortical activation to anger-inducing stimuli. *Int. J. Psychophysiol.* 66, 154–160. doi: 10.1016/j.ijpsycho.2007.03.020

Hinterberger, T., Neumann, N., Pham, M., Kübler, A., Grether, A., Hofmayer, N., et al. (2004). A multimodal brain-based feedback and communication system. *Exp. Brain Res.* 154, 521–526. doi: 10.1007/s00221-003-1690-3

Hirshfield, L., Girouard, A., Solovey, E., Jacob, R., Sassaroli, A., Tong, Y., et al. (2007). "Human-computer interaction and brain measurement using functional near-infrared spectroscopy," in *Proceedings of the ACM UIST'07 Symposium on User Interface Software and Technology* (Newport, RI: ACM Press). Available online at: https://uist.acm.org/archive/adjunct/2007/pdf/posters/p55-hirshfield.pdf

Hong, K.-S., Ghafoor, U., and Khan, M. J. (2020). Brain-machine interfaces using functional near-infrared spectroscopy: a review. *Artif. Life Robot.* 25, 204–218. doi: 10.1007/s10015-020-00592-9

Karran, A. J., Fairclough, S. H., and Gilleade, K. (2015). A framework for psychophysiological classification within a cultural heritage context using interest. *ACM Trans. Comput. Hum. Interact.* 21, 1–19. doi: 10.1145/2687925

Keynan, J. N., Cohen, A., Jackont, G., Green, N., Goldway, N., Davidov, A., et al. (2019). Electrical fingerprint of the amygdala guides neurofeedback training for stress resilience. *Nat. Hum. Behav.* 3, 63–73. doi: 10.1038/s41562-018-0484-3

Kohl, S. H., Mehler, D. M., Thibault, R. T., Konrad, K., Lührs, M., and Sorger, B. (2020). The potential of functional near-infrared spectroscopy-based neurofeedback–a systematic review and recommendations for best practice. *Front. Hum. Neurosci.* 14:594. doi: 10.3389/fnins.2020.00594

Lawrence, E. J., Su, L., Barker, G. J., Medford, N., Dalton, J., Williams, S. C., et al. (2014). Self-regulation of the anterior insula: reinforcement learning using real-time fMRI neurofeedback. *Neuroimage* 88, 113–124. doi: 10.1016/j.neuroimage.2013.10.069

Lee, M.-H., Fazli, S., Mehnert, J., and Lee, S.-W. (2014). "Hybrid brain-computer interface based on EEG and NIRS modalities," in *2014 International Winter Workshop on Brain-Computer Interface (BCI)* (Jeongsun-kun: IEEE), 1–2. doi: 10.1109/iww-BCI.2014.6782577

Li, K., Jiang, Y., Gong, Y., Zhao, W., Zhao, Z., Liu, X., et al. (2019). Functional Near-Infrared Spectroscopy (fNIRS) informed neurofeedback: regional-specific modulation of lateral orbitofrontal activation and cognitive flexibility. *bioRxiv* 511824. doi: 10.1101/511824

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., and Barrett, L. F. (2012). The brain basis of emotion: a meta-analytic review. *Behav. Brain Sci.* 35:121. doi: 10.1017/S0140525X11000446

Maior, H. A., Wilson, M. L., and Sharples, S. (2020). Workload alerts-using physiological measures of mental workload to provide feedback during tasks. *ACM Trans. Comput. Hum. Interac.* 25:9. doi: 10.1145/3173380

Mathiak, K. A., Alawi, E. M., Koush, Y., Dyck, M., Cordes, J. S., Gaber, T. J., et al. (2015). Social reward improves the voluntary control over localized brain activity in fMRI-based neurofeedback training. *Front. Behav. Neurosci.* 9:136. doi: 10.3389/fnbeh.2015.00136

Morinaga, K., Akiyoshi, J., Matsushita, H., Ichioka, S., Tanaka, Y., Tsuru, J., et al. (2007). Anticipatory anxiety-induced changes in human lateral prefrontal cortex activity. *Biol. Psychol.* 74, 34–38. doi: 10.1016/j.biopsycho.2006.06.005

Mühl, C., and Heylen, D. (2009). "Cross-modal elicitation of affective experience," in *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops* (Amsterdam: IEEE), 1–12. doi: 10.1109/ACII.2009.5349455

Naseer, N., and Hong, K.-S. (2015). fNIRS-based brain-computer interfaces: a review. *Front. Hum. Neurosci.* 9:3. doi: 10.3389/fnhum.2015.00003

Peck, E. M., Afergan, D., Yuksel, B. F., Lalooses, F., and Jacob, R. J. (2014). "Using fNIRS to measure mental workload in the real world," in *Advances in Physiological Computing*, eds S. H. Fairclough, and K. Gilleade (London: Springer), 117–139. doi: 10.1007/978-1-4471-6392-3_6

Ros, T., Enriquez-Geppert, S., Zotev, V., Young, K. D., Wood, G., Whitfield-Gabrieli, S., et al. (2020). Consensus on the reporting and experimental design of clinical and cognitive-behavioural neurofeedback studies (CRED-nf checklist). *Brain* 143, 1674–1685. doi: 10.1093/brain/awaa009

Ros, T., J Baars, B., Lanius, R. A., and Vuilleumier, P. (2014). Tuning pathological brain oscillations with neurofeedback: a systems neuroscience framework. *Front. Hum. Neurosci.* 8:1008. doi: 10.3389/fnhum.2014.01008

Rosenfeld, J. P., Cha, G., Blair, T., and Gotlib, I. H. (1995). Operant (biofeedback) control of left-right frontal alpha power differences: potential neurotherapy for affective disorders. *Biofeedb. Selfregul.* 20, 241–258. doi: 10.1007/BF01474516

Ruocco, A. C., Rodrigo, A. H., Lam, J., Di Domenico, S., Graves, B., and Ayaz, H. (2014). A problem-solving task specialized for functional neuroimaging: validation of the Scarborough adaptation of the tower of London (s-TOL) using near-infrared spectroscopy. *Front. Hum. Neurosci.* 8:185. doi: 10.3389/fnhum.2014.00185

Sakatani, K., Takemoto, N., Tsujii, T., Yanagisawa, K., and Tsunashima, H. (2013). NIRS-based neurofeedback learning systems for controlling activity of the prefrontal cortex. *Adv. Exp. Med. Biol.* 789, 449–454. doi: 10.1007/978-1-4614-7411-1_60

Sarkheil, P., Zilverstand, A., Kilian-Hütten, N., Schneider, F., Goebel, R., and Mathiak, K. (2015). fMRI feedback enhances emotion regulation as evidenced by a reduced amygdala response. *Behav. Brain Res.* 281, 326–332. doi: 10.1016/j.bbr.2014.11.027

Sitaram, R., Caria, A., and Birbaumer, N. (2009). Hemodynamic brain-computer interfaces for communication and rehabilitation. *Neural Netw.* 22, 1320–1328. doi: 10.1016/j.neunet.2009.05.009

Sitaram, R., Ros, T., Stoeckel, L., Haller, S., Scharnowski, F., Lewis-Peacock, J., et al. (2017). Closed-loop brain training: the science of neurofeedback. *Nat. Rev. Neurosci.* 18, 86–100. doi: 10.1038/nrn.2016.164

Solovey, E. T., Girouard, A., Chauncey, K., Hirshfield, L. M., Sassaroli, A., Zheng, F., et al. (2009). "Using fNIRS brain sensing in realistic HCI settings: experiments and guidelines," in *Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology* (Victoria, BC), 157–166. doi: 10.1145/1622176.1622207

Tachtsidis, I., and Scholkmann, F. (2016). False positives and false negatives in functional near-infrared spectroscopy: issues, challenges, and the way forward. *Neurophotonics* 3:031405. doi: 10.1117/1.NPh.3.3.031405

Tanaka, S. C., Samejima, K., Okada, G., Ueda, K., Okamoto, Y., Yamawaki, S., et al. (2006). Brain mechanism of reward prediction under predictable and unpredictable environmental dynamics. *Neural Netw.* 19, 1233–1241. doi: 10.1016/j.neunet.2006.05.039

Trambaiolli, L. R., Biazoli, C. E., Cravo, A. M., Falk, T. H., and Sato, J. R. (2018). Functional near-infrared spectroscopy-based affective neurofeedback: feedback effect, illiteracy phenomena, and whole-connectivity profiles. *Neurophotonics* 5:035009. doi: 10.1117/1.NPh.5.3.035009

Tuscan, L.-A., Herbert, J. D., Forman, E. M., Juarascio, A. S., Izzetoglu, M., and Schultheis, M. (2013). Exploring frontal asymmetry using functional near-infrared spectroscopy: a preliminary study of the effects of social anxiety during interaction and performance tasks. *Brain Imaging Behav.* 7, 140–153. doi: 10.1007/s11682-012-9206-z

Yamin, H. G., Gazit, T., Tchemodanov, N., Raz, G., Jackont, G., Charles, F., et al. (2017). Depth electrode neurofeedback with a virtual reality interface. *Brain Comput. Interfaces* 4, 201–213. doi: 10.1080/2326263X.2017.1338008

Zilverstand, A., Sorger, B., Slaats-Willemse, D., Kan, C. C., Goebel, R., and Buitelaar, J. K. (2017). fMRI neurofeedback training for increasing anterior cingulate cortex activation in adult attention deficit hyperactivity disorder. An exploratory randomized, single-blinded study. *PloS ONE* 12:e0170795. doi: 10.1371/journal.pone.0170795

Zohdi, H., Scholkmann, F., and Wolf, U. (2020). Frontal cerebral oxygenation asymmetry: intersubject variability and dependence on systemic physiology, season, and time of day. *Neurophotonics* 7, 1–19. doi: 10.1117/1.NPh.7.2.025006

Zotev, V., Yuan, H., Misaki, M., Phillips, R., Young, K. D., Feldner, M. T., et al. (2016). Correlation between amygdala BOLD activity and frontal EEG asymmetry during real-time fMRI neurofeedback training in patients with depression. *Neuroimage Clin.* 11, 224–238. doi: 10.1016/j.nicl.2016.02.003

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 06 July 2021
doi: 10.3389/fnins.2021.647844

A Novel Cognition-Guided Neurofeedback BCI Dataset on Nicotine Addiction

Junjie Bu[1,2†], Chang Liu[1†], Huixing Gou[1], Hefan Gan[1], Yan Cheng[3], Mengyuan Liu[3], Rui Ni[4], Zhen Liang[2], Guanbao Cui[5], Ginger Qinghong Zeng[1] and Xiaochu Zhang[1,3,5,6,7*]*

[1] Department of Radiology, The First Affiliated Hospital of USTC, Hefei National Laboratory for Physical Sciences at the Microscale and School of Life Science, Division of Life Science and Medicine, University of Science and Technology of China, Hefei, China, [2] Department of Intelligent Medical Engineering, School of Biomedical Engineering, Anhui Medical University, Hefei, China, [3] Department of Psychology, School of Humanities and Social Science, University of Science and Technology of China, Hefei, China, [4] Department of Life Sciences, Imperial College London, London, United Kingdom, [5] Institute of Advanced Technology, University of Science and Technology of China, Hefei, China, [6] Hefei Medical Research Center on Alcohol Addiction, Anhui Mental Health Center, Hefei, China, [7] Academy of Psychology and Behavior, Tianjin Normal University, Tianjin, China

OPEN ACCESS

Edited by:
Andrea Kübler,
Julius Maximilian University of Würzburg, Germany

Reviewed by:
Hui Zheng,
Zhejiang Normal University, China
Aleksandra Dagmara Kawala-Sterniuk,
Opole University of Technology, Poland

***Correspondence:**
Junjie Bu
bujunjie@ahmu.edu.cn
Xiaochu Zhang
zxcustc@ustc.edu.cn

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Neural Technology, a section of the journal Frontiers in Neuroscience

Received: 30 December 2020
Accepted: 27 May 2021
Published: 06 July 2021

Citation:
Bu J, Liu C, Gou H, Gan H, Cheng Y, Liu M, Ni R, Liang Z, Cui G, Zeng GQ and Zhang X (2021) A Novel Cognition-Guided Neurofeedback BCI Dataset on Nicotine Addiction. Front. Neurosci. 15:647844. doi: 10.3389/fnins.2021.647844

Compared with the traditional neurofeedback paradigm, the cognition-guided neurofeedback brain–computer interface (BCI) is a novel paradigm with significant effect on nicotine addiction. However, the cognition-guided neurofeedback BCI dataset is extremely lacking at present. This paper provides a BCI dataset based on a novel cognition-guided neurofeedback on nicotine addiction. Twenty-eight participants are recruited and involved in two visits of neurofeedback training. This cognition-guided neurofeedback includes two phases: an offline classifier construction and a real-time neurofeedback training. The original electroencephalogram (EEG) raw data of two phases are provided and evaluated in this paper. The event-related potential (ERP) amplitude and channel waveform suggest that our BCI dataset is of good quality and consistency. During neurofeedback training, the participants' smoking cue reactivity patterns have a significant reduction. The mean accuracy of the multivariate pattern analysis (MVPA) classifier can reach approximately 70%. This novel cognition-guided neurofeedback BCI dataset can be used to develop comparisons with other neurofeedback systems and provide a reference for the development of other BCI algorithms and neurofeedback paradigms on addiction.

Keywords: brain-computer interface, cognition-guided neurofeedback, nicotine addiction, electro encephalogram, public dataset

INTRODUCTION

The brain–computer interface (BCI) is a hardware and software system integrated as the interface between the brain and the computer (Janapati et al., 2020). Considering time sensitivity and device portability, BCI system generally uses electroencephalogram (EEG), electrocorticogram (ECoG), functional magnetic resonance imaging (fMRI), functional near-infrared spectroscopy (fNIRS), magnetoencephalography (MEG), and positron emission tomography (PET) as imaging methods. Among them, EEG is the most widely used BCIs (Kwon et al., 2020). ECoG can collect purer signal

than EEG, but it is invasive (Korostenskaja et al., 2014). Although fMRI and fNIRS have high spatial resolution, their temporal resolution is low (Cui et al., 2011). Besides, MEG and PET require large and expensive equipment and are not suitable for large-scale applications (Stam, 2010).

Nowadays, BCI system includes ERP, steady-state visually evoked potential (SSVEP), motor imagery (MI), and emotional BCI. In recent years, BCI-related fields have developed. Neurofeedback, as the predecessor of BCI, is applied to improve normal cognitive abilities, such as the enhancement of attention and working memory (Hsueh et al., 2016). It is also more and more used in the field of psychiatry studies, such as depression (Trambaiolli et al., 2021), anxiety (Gadea et al., 2020), addiction (Posson, 2019), etc.

Previous addiction-related neurofeedback datasets were usually based on fixed EEG or fMRI signals. Traditional EEG-based neurofeedback usually focuses on arousal/anxiety symptom in drug addiction by regulating α, α/θ, and sensorimotor rhythm (SMR)/β signals (Sokhadze et al., 2008). Real-time fMRI neurofeedback usually focuses on the activation of anterior cingulate cortex (ACC) or functional connectivity of bilateral ACC, medial prefrontal cortex (mPFC), orbitofrontal cortex (OFC), etc. (Martz et al., 2020). These datasets mentioned above have the following problems:

- No cognition task is performed to disclose the relationship between signals and behaviors/cognition components.
- Individual differences are not considered as all the subjects use a single and fixed signal to regulate the addiction behaviors.
- The efficacy of the traditional neurofeedback urges to be further improved in the treatment of addiction, as it is only rated as "probably efficacious."

Therefore, it is of vital importance to propose a novel neurofeedback paradigm.

We attempt to resolve the shortcomings of traditional neurofeedback by proposing the EEG cognition-guided neurofeedback based on a cue reactivity model. According to this model, when smokers are presented with visual, taste, or tactile cues related to smoking, they will have significant cue reactivity (physiological and physical reaction). Similar reactivity is found in healthy control subjects to non-drug evocative stimuli (Garavan et al., 2000). Cue reactivity leads to impulsive behavior in drug-seeking behavior as well as relapse (Chiamulera, 2005). The cue reactivity task is usually used to induce cue reactivity of participants. In our study, we used a smoking cue reactivity task to induce specific cue reactivity. Cue reactivity has multiple EEG features including both time- (e.g., P300, slow positive wave) and frequency-domain (e.g., alpha oscillation) features. Compared with the single signal, combing with multiple features (both time and frequency domain) using multivariate pattern analysis can better enhance sensitivity of detecting a particular brain activity pattern (Littel et al., 2012; Campanella et al., 2014). Our cognition-guided neurofeedback regulated the multiple signals induced by the specific cue reactivity. Therefore, the cognition-guided neurofeedback process included a specific addiction-related cognitive model (cue reactivity model). Based on this model, we performed the cognitive task (cue reactivity task) to obtain the specific addiction-related brain activities represented by multiple EEG features. In addition, this paradigm achieved a good intervention effect for smokers: the number of cigarettes consumed per day decreased 30.6, 38.2, and 27.4% compared with the baseline (pre-neurofeedback) at 1 week, 1 month, and 4 months follow-up (Bu et al., 2019). Our BCI dataset is based on a novel neurofeedback paradigm, which is closed loop, individualized, and MVPA based. Previous researchers have proposed a framework for cognitive neurofeedback in food cravings (Sokunbi, 2014; Ihssen et al., 2017) and cocaine addiction (Kirschner et al., 2018), but most of these studies are based on fMRI.

According to the cue reactivity model, smoking behavior enhances the conditional value of smoking cues, which is specifically reflected at two levels (**Supplementary Figure 1**):

- Bottom–up automatic processing to activate attentional functions.
- Top–down modulation of sensory inputting and motor controlling from cortical area.

According to conditioned reflex learning theory (Rees and Heather, 1995; Versace et al., 2017), the craving for cigarettes induced by smoking cues may partly relate to the conditioned reflex established by learning to associate smoking-related cues with smoking behavior (Karelitz, 2020). The reinforcement of smoking behavior also enhances the conditioned value of smoking cues, which automatically activates the attention function from the bottom to up. Besides, previous fMRI study has reported the regulation of cue-induced cravings from the top to down (Li et al., 2013).

Different from other neutral neurofeedback (for example, the height of the column is used for neurofeedback visual signals) (Zotev et al., 2014), we used an adaptive closed-loop method to develop our cognition-guided neurofeedback in the process of self-regulation training (Bu et al., 2019). The essence of this method is that when the participants' task performance decreased, the program would arouse the participants' attention and alertness by increasing the difficulty of the current task (deBettencourt et al., 2015). We established a mapping relationship between the probabilistic score and the 11 pictures (the probabilistic score was positively correlated with the desiring rating of the picture). Participants' brain activity state toward smoking cues was reflected as smoking-related pictures with different craving levels. At the same time, the picture would affect the brain activity pattern in the next trial, which was mapped to the corresponding smoking-related picture. In other words, we amplified the consequence of neurofeedback training: rewarding successful downregulation by reducing difficulty and punishing unsuccessful downregulation by increasing difficulty to activate participants' self-monitoring ability. By this way, an adaptable closed-loop effect was formed, which is one of the characteristics of our cognition-guided neurofeedback.

Neurofeedback intervention methods may produce different effects in different participants. Different visits of the same

participant will also lead to different EEG signals (Gruzelier, 2014). Considering the individual differences of neurofeedback intervention and changeable craving of a specific participant, our neurofeedback paradigm built an individual model for each visit of each participant, which eliminated individual differences to a great extent. Each subject had exclusive classifier, and the regulated EEG signals (time and frequency domain) were obtained from his/her own cue reactivity task. The model used by the subject was reconstructed in the current cycle. Traditional neurofeedback regulated fixed signal, which may be not suitable in some subjects. Individual classification is one of the characteristics of our cognition-guided neurofeedback.

Traditional dependent neurofeedback usually includes single signal. There are also researches about neurofeedback based on network features, such as the algorithm based on common spatial pattern (CSP) and local characteristic-scale decomposition (LCD) developed by Ai et al. (2019). Addiction is a complex pattern of brain activity, rather than just being related to a certain electrical brain signal, such as P300 (Littel et al., 2012) and α power (Cui et al., 2013). The development of a classifier with more features or variables is particularly important. MVPA classifier is used to extract features of different dimensions from cue reactivity task (Zafar et al., 2017). This algorithm based on machine learning can detect the brain reactivity patterns in response to smoking and neutral stimulus more sensitively (Sitaram et al., 2017).

Our novel paradigm has a good performance and shows significant short- and long-term effects (Bu et al., 2019), but so far, there is a lack of novel neurofeedback-related datasets. Sharing these datasets online can facilitate comparison with other neurofeedback datasets, promote parameter optimization process, and help optimize the BCI algorithm. Single data may be limited. More datasets shared online can also be applied to big data models to get better results. In addition, BCI datasets are also instructive for BCI hardware development. In the field of psychiatry, more datasets on clinical nicotine addiction are needed to promote the study of addiction mechanisms, especially smoking cue reactivity mechanisms.

In this paper, we provide a novel dataset based on a cognition-guided, closed-loop, and individualized neurofeedback, which is based on MVPA classifier. The dataset includes EEG data of two phases: the cue reactivity task and the real-time neurofeedback training. In addition, we evaluated the quality of EEG data in our dataset by ERP and topographic map analysis. A linear correlation is used to indicate the trend of probabilistic score in neurofeedback training. Prediction accuracies of each participant are used to evaluate the classification power. The publication of novel cognition-guided neurofeedback dataset is of vital importance to the development of this field.

MATERIALS AND METHODS

Participants

In this study, 28 participants (male; mean age, 23.7 years) were recruited through online advertisements and posters by the criteria listed below. The score of Fagerstrom Test for Nicotine Dependence (FTND) of the 28 subjects were 4.6 ± 1.9 (mean ± SD, **Table 1**). Since there are only a few female smokers (2.7%) in China, we recruited only male participants for this experiment. This study was approved by the Human Ethics Committee of the University of Science and Technology of China (USTC). According our records, no participant reported uncomfortable feelings after neurofeedback training (Bu et al., 2019).

Selection criteria were as follows:

- Smoking 10 cigarettes or more per day for at least 2 years
- Right-handedness
- 18–40 years old
- Normal or corrected-to-normal vision
- Normal mental and physical health condition assessed by the Mini-International Neuropsychiatric Interview (MINI).

Exclusion criteria were as follows:

- Chronic neurological, psychiatric, or medical disease
- Taking any drugs in the past 3 months
- Unable to perform EEG for any reasons.

Experimental Design

We developed a novel cognition-guided neurofeedback paradigm on nicotine addiction. Participants were involved in two continuous visits over 2 days. In each visit, participants experienced a two-phase procedure shown in **Figure 1C**: offline classifier construction and real-time EEG neurofeedback training.

Before the formal experiment, we prepared 330 pictures (210 × 180 pixels), including 150 smoking-related pictures (e.g., holding a cigarette in hand), 150 paired neutral pictures (e.g., holding a pencil in hand), and 30 pictures of animals (e.g., cat). The selection of these pictures mainly referred to the previous work on addiction of our laboratory (Zhang, 2011). To eliminate the possible impact of visual information, these smoking pictures and neutral pictures were matched as much as possible in terms of visual information such as background color, brightness, and object orientation. We recruited 20 participants to evaluate the craving degree of these pictures. The picture evaluation procedure is shown in **Figure 2A**. Eleven pictures were selected for real-time EEG neurofeedback training from 150 smoking-related pictures. They were listed in ascending order of the craving score given by the participants (**Figure 2B**).

TABLE 1 | Demographic information of 28 participants.

Characteristic	Value
Age (years)	23.7 (3.8)
Education (years)	14.8 (2.5)
Cigarettes (day)	14.1 (4.5)
Cigarette use (years)	7.1 (3.9)
FTND score	4.6 (1.9)

Values were mean and values in parentheses were 1 standard deviation (SD).

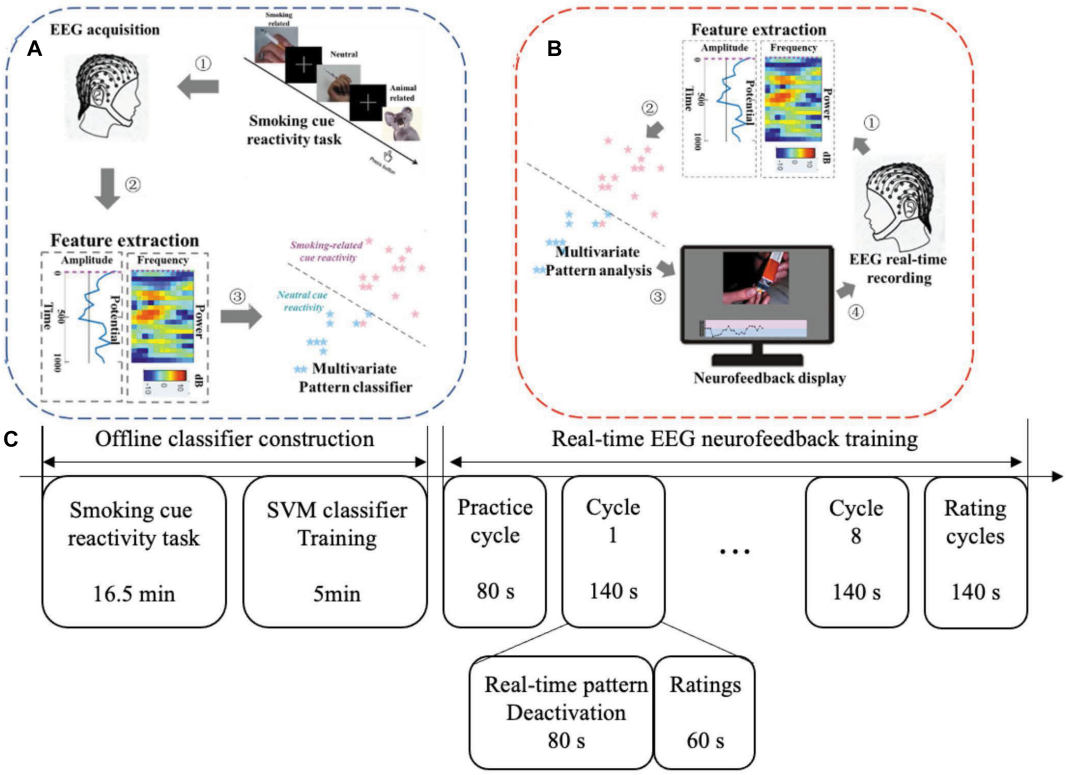


FIGURE 1 | Two phases for one neurofeedback visit: offline model construction and real-time neurofeedback training. **(A)** Offline classifier construction phase. **(B)** Real-time EEG neurofeedback training phase. **(C)** Experimental protocol in one visit. NF, neurofeedback.

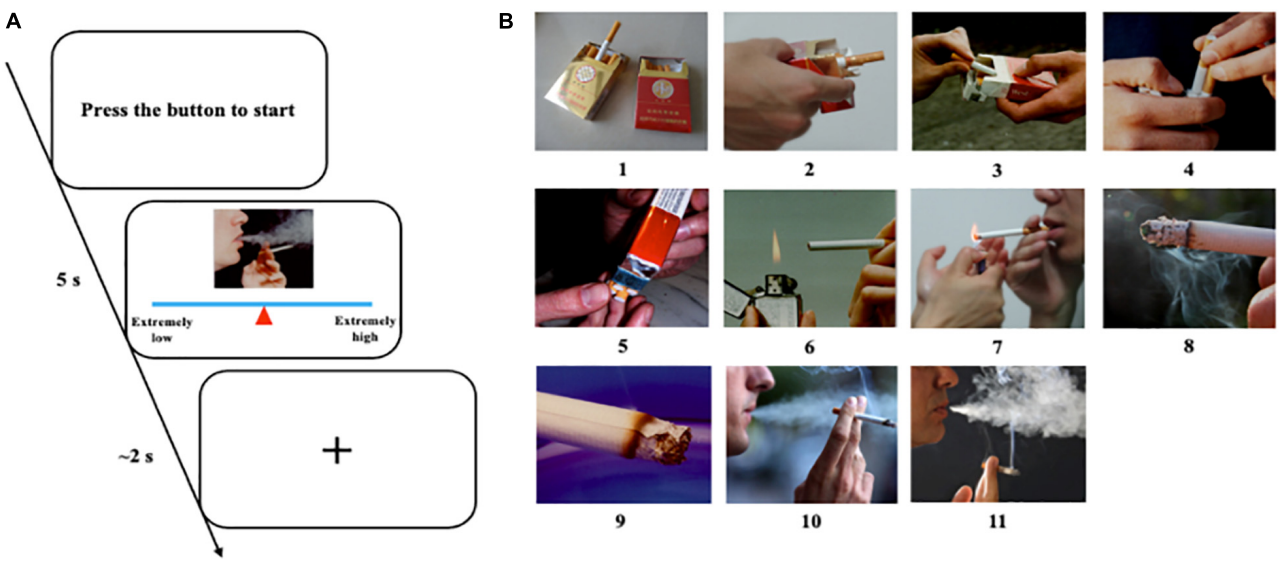


FIGURE 2 | Evaluation of nicotine related cues. **(A)** Picture evaluation protocol and procedure. Participants pressed the button to start a new trial. They had 5 s to move the mouse to change the position of the triangle on the line to give an appropriate score according to their craving to this picture. Then, they had about 2 s to rest and wait for the next trial. **(B)** Eleven selected pictures listed in ascending order of craving score.

Offline Classifier Construction

Offline classifier construction consisted of two parts: the smoking cue reactivity task and the support vector machines (SVMs) classifier training (**Figure 1A**). EEG data collected from cue reactivity task were used to train the classifier.

The smoking cue reactivity task was block design. Three smoking and neutral blocks each were assessed in a pseudo-randomized order (neutral, smoking, smoking, neutral, smoking, and neutral). There were 55 trials in each block, and participants were requested to focus on the pictures as much as possible. Among them, 50 trials were smoking or neutral cues. Another five randomly distributed trials were animal cues, which were used to improve the concentration of participants on the task. They were asked to press the space button of the keyboard quickly and accurately when seeing animal cues (the main objects are animals, such as cats and dogs) and press the space button of the keyboard quickly and accurately. At the same time, EEG data were collected during each block. In the block interval, participants had 90 s of rest. The smoking cue reactivity program was written by Psychophysics Toolbox[1] for MATLAB (version 2016a, MathWorks Inc., Natick, MA, United States).

EEG data were processed offline in order to remove noise using EEGLAB (more details and parameters will be mentioned in *Data processing*). Permutation test was performed to extract features in time and frequency domain from the processed data ($\alpha = 0.05$). Specifically, we compared EEG data collected from smoking and neutral blocks to obtain significantly different features in time (amplitude) and frequency (power) domain. These statistically significant features were transferred to cluster features via cluster-based statistic (maximum cluster-level mass) and imported into the classifier in the form of a column vector to train a personalized SVM classifier to recognize brain activity patterns in response to smoking stimuli. This step was calculated using the MATLAB function fitcsvm (Bu et al., 2019). Sixty channels (64 standard channels except for CB1, CB2, HEOG, and VEOG) were used to construct the model. In order to evaluate the classification effectiveness of the classifier, we used 20% of the trials to calculate the prediction accuracy in each cross-validation cycle. This step repeated five times for each participant.

Real-Time EEG Neurofeedback Training

The neurofeedback training phase comprised eight same cycles in total, including 40 trials (80 s) and a rating (60 s) in each phase. Before the first trial, we carried out a practice cycle to make sure that participants understand the experimental requirements (**Figure 3**).

The real-time neurofeedback signal of each trial was updated every 2 s, including 1 s of real-time EEG acquisition and classification each. Online preprocessing (the same with offline processing mentioned above) was performed to remove noise. Processed EEG raw data were put into the classifier. Based on the SVM classifier, a probabilistic score (range, 0–1) was returned to characterize the similarity probability between the current brain activity and the brain activity pattern of smoking measured in smoking cue reactivity task. The probabilistic score was presented to participants in real time as a moving point in the feedback line at the bottom of the screen (**Figure 1B**). In order to prevent drastic changes of the line, the value of each point was obtained by averaging the value of the current point and the previous 2 points, and the first 2 points in the line were fixed 0.5 (no previous points to be averaged). At the same time, based on this value, the corresponding type of picture was displayed on the top of the screen according to adaptive closed-loop design. A higher probabilistic score corresponded to a picture with a higher craving level. This closed-loop neurofeedback program was written by Psychophysics Toolbox for MATLAB.

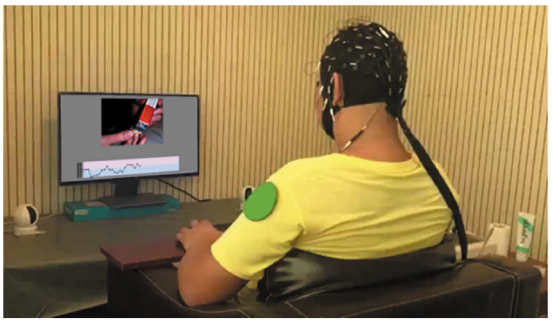


FIGURE 3 | One scene of a participant in the phase of neurofeedback regulation training. A participant is wearing EEG cap and watching neurofeedback display. A camera to monitor the status of the participant.

In this phase, participants were asked to downregulate the line repeatedly and continuously while fixating the picture. If one strategy failed to downgrade the line well, they needed to change the strategy until they found the most effective one. After finishing training and rating cycles (**Figure 1C**), participants were asked to record 10 mental strategies that may effectively downregulate the neurofeedback signals. To improve the initiative of participants, we informed them about the rewards they can get for completing the experiment in advance. Participants who completed eight cycles would receive a reward of up to 140 RMB according to their performance (proportional to mean probabilistic score). After completing a cycle, participants would have a rest for 1 min.

Data Acquisition

EEG raw data were collected using a SynAmps RT amplifier (NeuroScan, Inc., Sterling, VA, United States), and an electrode cap was with 64 Ag/AgCl electrodes located according to international 10–20 system. Additionally, the left (M1) and right mastoids (M2) were also recorded. The impedance of all electrodes was kept under 5 kΩ based on the reference electrode attached to the tip of the nose. The EOG was recorded by VEOG (above and below the left eye) and HEOG (lateral to the outer canthi of both eyes) using four electrodes (VEOL, VEOU, HEOL, and HEOR). In order to reduce the influence of electromagnetic interference on EEG signals, we grounded the AFz electrode. The sampling rate of the EEG raw data was 250 Hz. In neurofeedback sessions, EEG raw signals were collected with the same parameters and sent to MATLAB program using NeuroScan Access SDK (NeuroScan, Inc., Sterling, VA, United States).

[1] http://psychtoolbox.org/

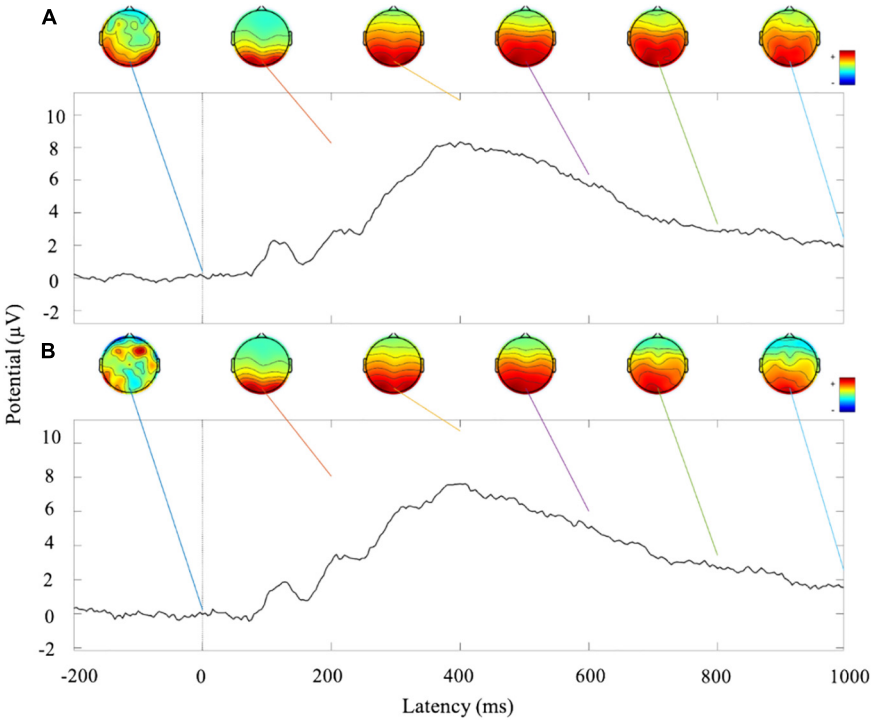


FIGURE 4 | Analysis of ERP amplitude and topographic map of cue reactivity task. **(A)** ERP waveform and topographic maps of the first visit of cue reactivity task phase. **(B)** ERP waveform and topographic maps of the second visit of cue reactivity task phase.

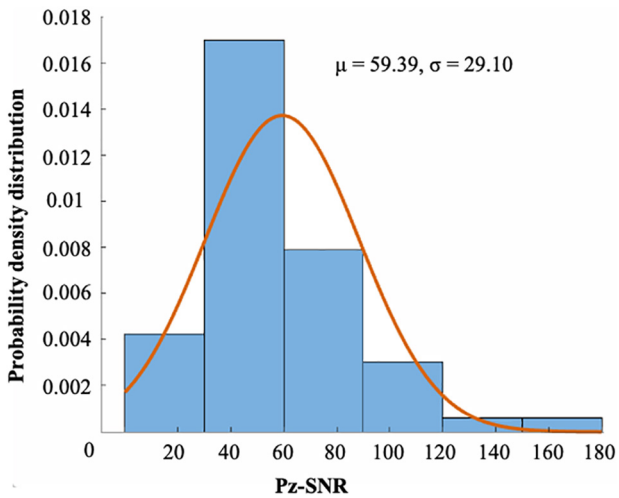


FIGURE 5 | The distribution of the SNR of Pz channel (μ = 59.39, σ = 29.10). There were 55 runs of 28 subjects. There was no significant difference between the distribution and the normal distribution by KS test (p = 0.29).

Data Processing

The EEG raw data processing was conducted by EEGLAB toolbox (version 14_1_1b) for MATLAB. The preprocessing steps included high-pass filter (0.5 Hz), epoch (−200 to 1,000 ms, −200 to 0 ms as prestimulus interval to conduct baseline correction), and blink artifacts (using a conventional recursive least squares algorithm). ERP analysis was also conducted in EEGLAB. Epochs containing the amplitude changes exceeding ±100 mV were rejected. The ERPs were grand averaged based on different types of stimulus (smoking and neutral) across participants. The power characteristics of the time–frequency domain were mainly obtained based on the wavelet analysis algorithm. The frequency of the EEG data was divided into five ranges: alpha, low beta, high beta, low gamma, and high gamma waves.

Data Evaluation

At the offline classifier construction phase of the cue reactivity task, EEG raw data were collected, and the signal-to-noise ratio (SNR) of Pz channel was calculated to verify the quality of the data (Hu et al., 2010). EEG epochs from −200 to 1,000 ms were preprocessed and averaged to calculate ERP waveforms separately for two visits. Besides, topographic maps were shown every 200 ms in the time window of 0–1,000 ms. Processed EEG data were imported to the classifier, and fivefold cross-validation accuracy was used to calculate its prediction accuracy.

In neurofeedback training phase, 1-s EEG data were preprocessed and shown at only channels located on the frontal midline. Topographic maps were shown every 200 ms in the time window of 0–1,000 ms. These 1-s EEG data were inputted to the personalized classifier model built in the offline phase, and a probabilistic score of smoking cue reactivity patterns was given. The score reflected the matching degree between the participant's current brain activity pattern and the pattern

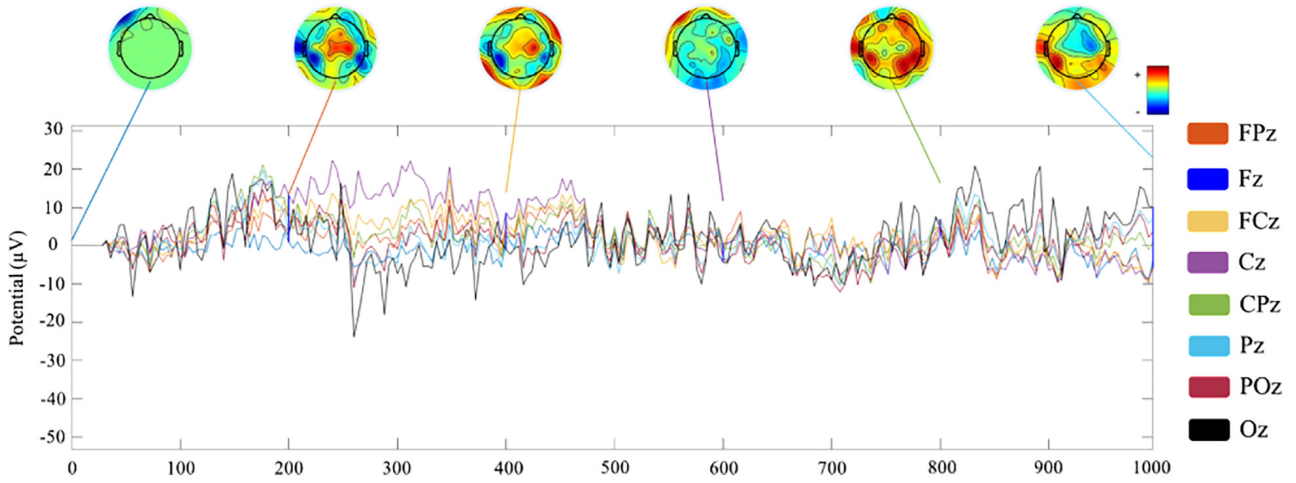


FIGURE 6 | Channel waveform and topographic map of real-time neurofeedback training phase.

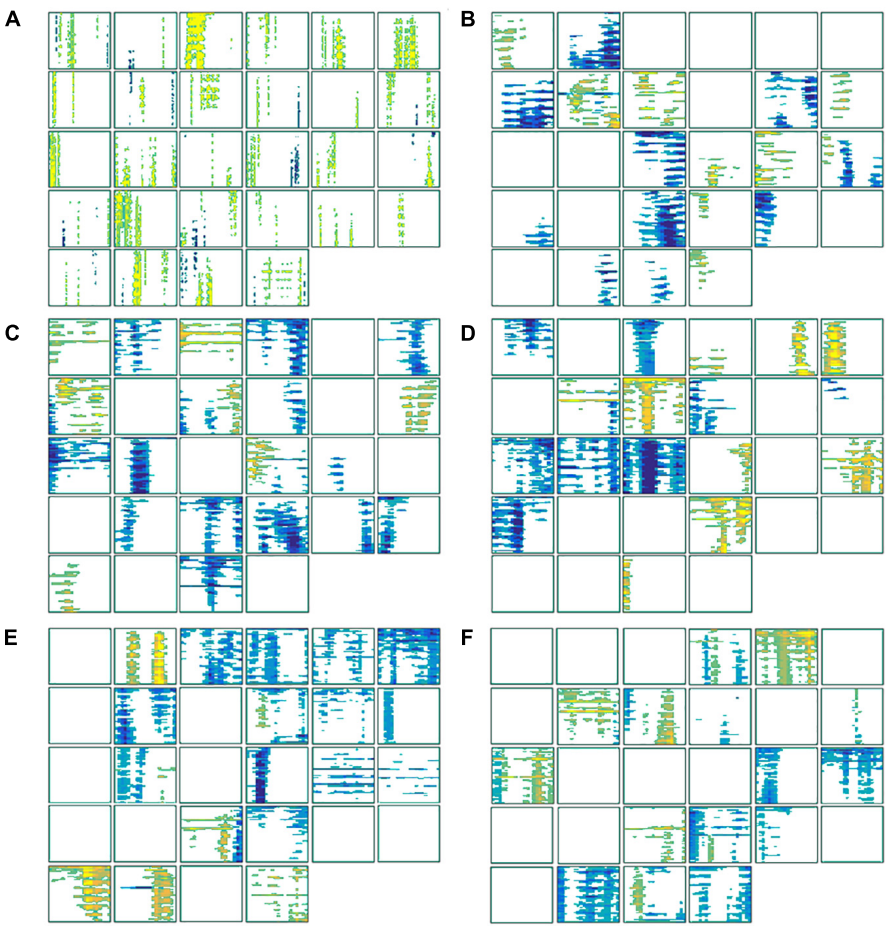


FIGURE 7 | The statistically significant features used to construct the classifier. **(A)** The features of each subject in the time domain; **(B)** the features in the alpha (8–13 Hz) frequency band; **(C)** the features in the low beta (14–20 Hz) frequency band; **(D)** the features in the high beta (21–30 Hz) frequency band; **(E)** the features in the low gamma (31–48 Hz) frequency band; **(F)** the features in the high-gamma (52–80 Hz) frequency band.

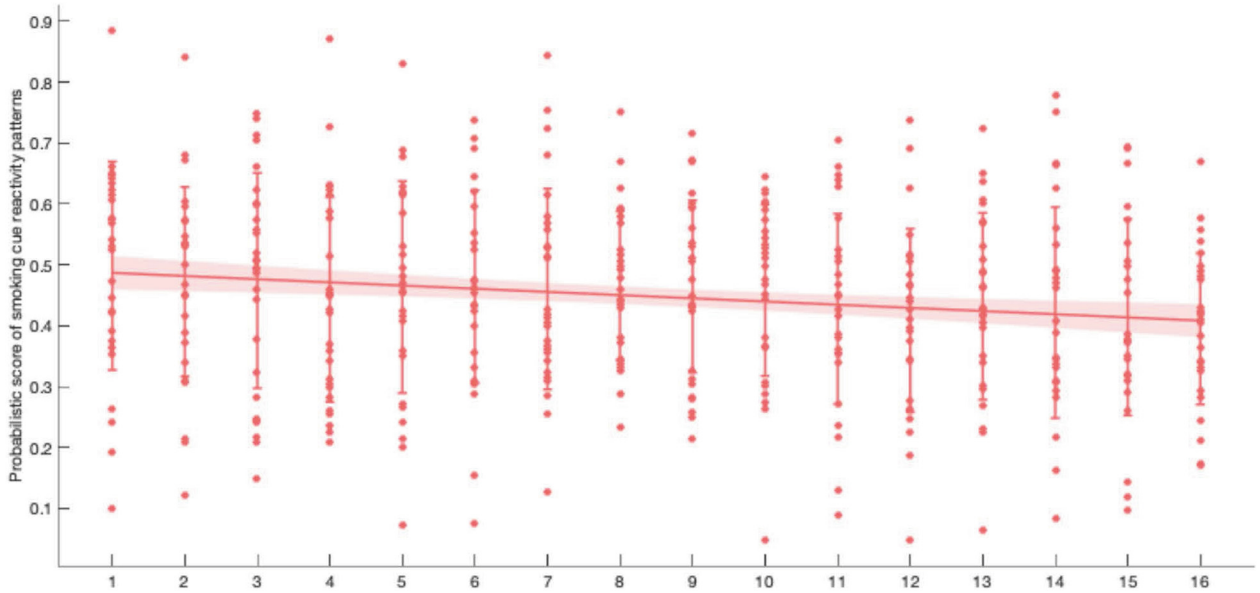


FIGURE 8 | Within two visit of neurofeedback learning, participants tried to reduce the probabilistic score (r = −0.1545, p = 0.0010). Error bar: SD; shaded area: 95% CI.

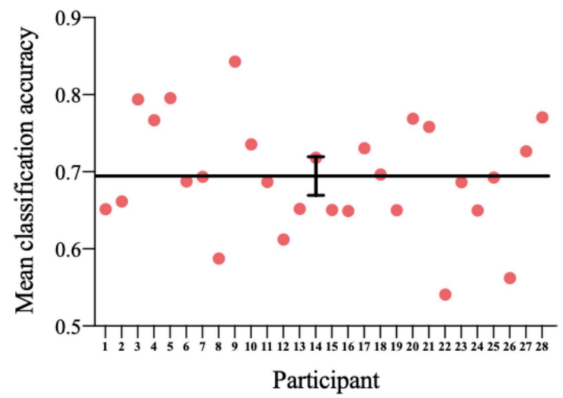


FIGURE 9 | The accuracy of the classifier for each participant. Horizontal line: average; error bar: SE.

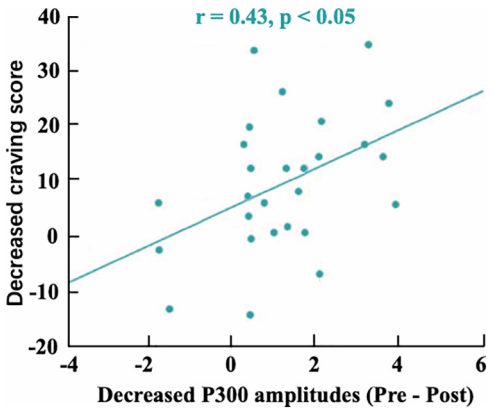


FIGURE 10 | The correlation between the mean decreased P300 amplitudes and decreased craving score (r = 0.43).

while viewing smoking cues in cue reactivity task. A high score indicated closely matched patterns. After training, the correlation between decreased P300 amplitudes (pre–post) and decreased craving score [score of tobacco craving questionnaire (TCQ), pre vs. post] was performed.

Statistical Analysis

Statistical analysis was conducted in Statistics and machine learning toolbox in MATLAB. The comparison of features was performed with a non-parametric permutation test. The correlation of probabilistic score and training cycles was analyzed using Pearson's correlation. The comparison between the prediction accuracy and chance level used one-sample Student's t-test. The normality test was based on Kolmogorov–Smirnov (KS) test. Fivefold cross-validation was used to calculate the classification accuracy of the classifier. The reported p-values were all two-tailed. The significant threshold α is 0.05.

RESULTS

EEG Data Evaluation

The database is 3.96 GB, including 2.44 GB EEG data of cue reactivity task, 1.52 GB EEG data of neurofeedback, 16.8 kB demographic data, and 1.81 kB channel location file. Twenty-eight subjects were included in this database. The cue reactivity data were divided into two files: "cue reactivity_1.zip" and

"cue reactivity_2.zip." Each subject had two runs, and each run contained six CNT files, which could be read in EEGLAB toolbox. The neurofeedback data were in "nf_eeg.zip." Each subject had two runs, and each run contained eight EEG MAT files (channel*time), which can be read in MATLAB. To facilitate subsequent data processing, electrode-position file and marker information were also provided. The channel location file named "chan62.zip," the baseline demographic and clinical characteristics named "baseline.mat," and trigger information file named "trigger.md" can be found in the root directory.

Figures 4A,B show the ERP amplitude and topographic map of cue reactivity task in the first and second visits. The data of the two visits have good consistency. The P300 component of ERP caused by addictive substance-related cues is usually a feature of substance use disorders, and the P300 amplitude is found to be related to the craving for smoking (Littel et al., 2012; Campanella et al., 2014). In our research, the P300 ERP component induced by smoking cues can be observed at approximately 300–550 ms. We evaluated the distribution of the Pz-SNR (**Figure 5**). KS test was performed to check the normality of distribution. There was no significant difference between the data distribution and the normal distribution ($p = 0.29$).

Channel waveform and topographic map of preprocessed 1-s real-time neurofeedback training data of eight channels (including FPz, Fz, FCz, Cz, CPz, Pz, POz, and Oz) are shown in different color to express brain electrical activity. We used the data of the 10th trial, the first cycle, the first visit (run1) of subject 1 (s1) as an example (**Figure 6**).

Feature Extraction Evaluation

These time (**Figure 7A**) and frequency domain (**Figures 7B–F**) features from 28 participants were used to construct the classifier. For each block diagram, the abscissa represents the time information (ms), and the ordinate represents electrode channel information (60 channels) respectively. The yellow and blue areas in the figure represent the characteristic signals that were statistically significant with smoking > non-smoking conditions and the characteristic signal that was statistically significant with smoking < non-smoking conditions ($\alpha = 0.05$), respectively.

Neurofeedback Performance Evaluation

Participants were trained for 16 cycles in two visits. In each trial of the neurofeedback cycle, the average score of 40 probabilistic scores was recorded as neurofeedback performance in the respective cycle. We performed a linear regression analysis on the probabilistic score and the training cycles. A significant negative correlation ($r = -0.1545$, $p = 0.0010$) was found, which showed that the match probability (the current brain activity patterns and the patterns when viewing smoking cues) decreased with training progressed (**Figure 8**). In other words, the difference in brain activity patterns between viewing the smoking cues and the neutral cue was smaller after neurofeedback training.

Classifier Power Evaluation

Each point in **Figure 9** represents fivefold cross-validation accuracy of one participant. The accuracy of the classifier varies between 0.5409 and 0.8427, with a mean accuracy of 0.6935. One-sample *t*-test showed a significant difference between the prediction accuracy and the chance level ($p = 4.4176 \times 10^{-14}$).

CONCLUSION AND DISCUSSION

In this paper, we developed a novel cognition-guided neurofeedback paradigm. Neurofeedback technology is an effective method to regulate brain signals and neuroplasticity, which involves brain networks of reward, control, and learning (Sitaram et al., 2017). Among them, dorsolateral prefrontal cortex (dlPFC) and posterior parietal cortex (PPC) in the control network will be activated during the execution of the strategy, and the learning network [mainly includes the dorsal striatum (DS)] is responsible for strategy learning in neurofeedback. The theoretical models of neurofeedback learning include operant (or instrumental) learning, motor learning, dual process theory, awareness theory, global workspace theory, and skill learning theory (Sitaram et al., 2017). These theoretical models are not mutually exclusive but compatible. Operant learning can be regarded as part of the dual process theory, which includes automatically process and controlled process (Enriquez-Geppert et al., 2017). Normally, approximately 15–30% of subjects do not respond to neurofeedback/BCI (Weber et al., 2011). Strategies are extremely important for neurofeedback. In the process of neurofeedback learning, strategies may promote or hinder neurofeedback learning, depending on the appropriate degree of scheduling cognitive resources in the process (Gaume et al., 2016; Davelaar, 2018).

In our implicit neurofeedback, the significant negative correlation ($r = -0.1545$, $p = 0.0010$) of the probabilistic score and training cycles demonstrates good neurofeedback learning effect. It also shows that the brain activity pattern in response to smoking stimuli represented by EEG signals can be successfully downregulated after two neurofeedback trainings. This is similar to previous rt-fMRI neurofeedback studies based on visual perception and attention brain activity patterns (deBettencourt et al., 2015; Amano et al., 2016). Our research also shows that the changes in brain patterns influenced addictive behaviors. The subjects' TCQ craving score and P300 amplitude decreased significantly after neurofeedback training (Bu et al., 2019), and there was a positive correlation between them (**Figure 10**). The results revealed the relationship between neural signals and behavioral indicators.

In particular, previous neurofeedback studies based on the regulation of brain activity patterns focused on training normal subjects to improve their cognitive abilities. The results of our study further extended these findings to patients with mental illness (smoking addicts), which also had the potential to regulate brain activity patterns. Our dataset provided important support for the extension of neurofeedback training of multivariable brain activity pattern regulation to clinical research in the future.

This dataset is of high quality and good integrity. In order to reduce artifacts as much as possible and improve the SNR, we added simple cognitive activities while performing

the cue reactivity task to ensure that the collected data were brain electrical signals from the experiment rather than distractions. Additionally, in the cue reactivity task phase and the neurofeedback training phase, an intertrial interval (ITI) was used to eliminate the detention effect of the previous trial and gave participants a certain time to rest until they felt comfortable enough for the next trial. During the experiment, participants were instructed to blink and swallow as little as possible to obtain purer EEG raw data. As can be seen in **Figure 4**, ERP waveforms and topographic maps have strong consistency and repeatability between visits. In addition, we checked the strategies reported by the participants and the surveillance video during neurofeedback and found that no subjects used the strategies that were not allowed, such as not watching the screen. The decline in smoking cue reactivity patterns and the relatively high model prediction accuracy provided a guarantee for the intervention effect and method reliability of neurofeedback BCI. Nevertheless, the dataset also had some limitations. Considering that the prevalence among female smokers in China was very low (2.7%), we only recruited male participants. The dataset we present is the first cognition-guided neurofeedback BCI dataset on nicotine addiction to our knowledge. Participants completing two separate visits of training showed improvement in smoking cue reactivity patterns. High-quality EEG raw data are provided, and the classifier is evaluated and proven to have relatively high classification accuracy. Our two-visit, cognition-guided neurofeedback BCI can be compared with other neurofeedback paradigms to develop new neurofeedback systems.

Our previous research extracted the characteristics of the EEG signal in the time and frequency domains for SVM modeling. As far as we know, the cue reactivity task is commonly studied in substance addiction (Hardy et al., 2017). The dataset we provided contains the participants' brain electrical activity exposed to both smoking-related cues and neutral cues. There are other methods to analyze the dataset, such as the microstate (Michel, 2018) and source analysis (Zhou et al., 2019). These methods can be implemented on this dataset to discover mechanisms of nicotine addiction. The EEG data provided in this study can be used to verify other existing models or optimize parameters. In addition, the dataset can be used to development other ERP-based BCI algorithms.

In future research, BCI datasets can still be improved from the following aspects. First, a higher number of channels and sampling rate can be applied to obtain higher BCI bandwidth. Second, the extraction of feature can be improved. Different network connection of participants toward smoking and neutral cues can be incorporated into classification features to obtain a higher degree of discrimination (Ai et al., 2019). Third, SVM classifier was used in this study, and a relatively high classification accuracy rate was obtained. Subsequent studies can use other machine learning method to obtain higher model prediction rates, such as logistic regression and decision tree. For example, a feature extraction method was developed based on autoregression and used random forest classifiers to identify the EEG features of patients with epilepsy with a best accuracy of 97.352% (Zhang et al., 2017). Fourth, this study conducted two visits of neurofeedback training, which was mainly based on skill learning theory (Hinterberger et al., 2005; Koralek et al., 2012) and previous MRI research (Young et al., 2017a,b). In future BCI studies, longer training periods can be used to evaluate the impact of training time on neurofeedback effects. Finally, this study used a simple floating line as the visual form of neurofeedback. Nowadays, BCI is moving in the direction of gamification (de Castro-Cros et al., 2020), which is also the direction for future BCI datasets.

In conclusion, our novel neurofeedback BCI dataset has a significant contribution to this field. We offer the community access to our EEG data from our BCI experiment. The neurofeedback protocol that we developed and applied is based on the long-term research on nicotine addiction of our group and will act as a reference for subsequent neurofeedback BCI research.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://github.com/Qingqingran/Cognition-guided-neurofeedback-BCI-dataset-on-nicotine-addiction/releases/tag/publish.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Human Ethics Committee of the University of Science and Technology of China. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

JB and XZ contributed to the experimental design. JB and GC contributed to the data collection. JB, CL, HGo, HGa, YC, ML, GQZ, and XZ contributed to the manuscript writing. CL, RN, and ZL contributed to the data processing. CL and HGo contributed to the data uploading. All listed authors made direct contributions to the article and approved it for publication.

FUNDING

This work was supported by grants from the National Key Basic Research Program (2018YFC0831101). The National Natural Science Foundation of China (31771221, 71942003, 61773360, 31800927, 31900766, 32000750, and 71874170). Major Project of Philosophy and Social Science Research, Ministry of Education of China (19JZD010), CAS-VPST Silk Road Science Fund 2021 (GLHZ202128), Collaborative Innovation Program of Hefei Science Center, CAS (2020HSC-CIP001), Anhui

Provincial Natural Science Foundation (2008085QH369), Major Project of Natural Science Foundation of the Anhui Higher Education Institutions (KJ2020ZD16), China Postdoctoral Science Foundation (2019TQ0312 and 2019M662203), Basic and Clinical Collaborative Research Improvement Project of Anhui Medical University (2020xkjT020), School Foundation of Anhui Medical University (2019xkj016), and Grants for Scientific Research of BSKY (XJ201907) from Anhui Medical University. A portion of the numerical calculations in this study was performed with the supercomputing system at the Supercomputing Centre of USTC.

ACKNOWLEDGMENTS

We sincerely acknowledge Wei Hong (WH) for their assistance with subject assessment.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.647844/full#supplementary-material

REFERENCES

Ai, Q., Chen, A., Chen, K., Liu, Q., Zhou, T., Xin, S., et al. (2019). Feature extraction of four-class motor imagery EEG signals based on functional brain network. *J. Neural Eng.* 16:026032. doi: 10.1088/1741-2552/ab0328

Amano, K., Shibata, K., Kawato, M., Sasaki, Y., and Watanabe, T. (2016). Learning to associate orientation with color in early visual areas by associative decoded fMRI neurofeedback. *Curr. Biol.* 26, 1861–1866. doi: 10.1016/j.cub.2016.05.014

Bu, J., Young, K. D., Hong, W., Ma, R., Song, H., Wang, Y., et al. (2019). Effect of deactivation of activity patterns related to smoking cue reactivity on nicotine addiction. *Brain* 142, 1827–1841. doi: 10.1093/brain/awz114

Campanella, S., Pogarell, O., and Boutros, N. (2014). Event-related potentials in substance use disorders: a narrative review based on articles from 1984 to 2012. *Clin. EEG Neurosci.* 45, 67–76. doi: 10.1177/1550059413495533

Chiamulera, C. (2005). Cue reactivity in nicotine and tobacco dependence: a "multiple-action" model of nicotine as a primary reinforcement and as an enhancer of the effects of smoking-associated stimuli. *Brain Res. Rev.* 48, 74–97. doi: 10.1016/j.brainresrev.2004.08.005

Cui, X., Bray, S., Bryant, D. M., Glover, G. H., and Reiss, A. L. (2011). A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *NeuroImage* 54, 2808–2821. doi: 10.1016/j.neuroimage.2010.10.069

Cui, Y., Versace, F., Engelmann, J. M., Minnix, J. A., Robinson, J. D., Lam, C. Y., et al. (2013). Alpha oscillations in response to affective and cigarette-related stimuli in smokers. *Nicotine Tob. Res.* 15, 917–924. doi: 10.1093/ntr/nts209

Davelaar, E. J. (2018). Mechanisms of Neurofeedback: A Computation-theoretic Approach. *Neuroscience* 378:175-188.

de Castro-Cros, M., Sebastian-Romagosa, M., Rodríguez-Serrano, J., Opisso, E., Ochoa, M., Ortner, R., et al. (2020). Effects of gamification in BCI functional rehabilitation. *Front. Neurosci.* 14:882. doi: 10.3389/fnins.2020.00882

deBettencourt, M. T., Cohen, J. D., Lee, R. F., Norman, K. A., and Turk-Browne, N. B. (2015). Closed-loop training of attention with real-time brain imaging. *Nat. Neurosci.* 18, 470–475. doi: 10.1038/nn.3940

Enriquez-Geppert, S., Huster, R. J., and Herrmann, C. S. (2017). EEG-neurofeedback as a tool to modulate cognition and behavior: a review tutorial. *Front. Hum. Neurosci.* 11:51. doi: 10.3389/fnhum.2017.00051

Gadea, M., Aliño, M., Hidalgo, V., Espert, R., and Salvador, A. (2020). Effects of a single session of SMR neurofeedback training on anxiety and cortisol levels. *Neurophysiol. Clin.* 50, 167–173. doi: 10.1016/j.neucli.2020.03.001

Garavan, H., Pankiewicz, J., Bloom, A., Cho, J.-K., Sperry, L., Ross, T. J., et al. (2000). Cue-induced cocaine craving: neuroanatomical specificity for drug users and drug stimuli. *Am. J. Psychiatry* 157, 1789–1798. doi: 10.1176/appi.ajp.157.11.1789

Gaume, A., Vialatte, A., Mora-Sánchez, A., Ramdani, C., and Vialatte, F. B. (2016). A psychoengineering paradigm for the neurocognitive mechanisms of biofeedback and neurofeedback. *Neurosci. Biobehav. Rev.* 68, 891–910. doi: 10.1016/j.neubiorev.2016.06.012

Gruzelier, J. H. (2014). EEG-neurofeedback for optimising performance. III: a review of methodological and theoretical considerations. *Neurosci. Biobehav. Rev.* 44, 159–182. doi: 10.1016/j.neubiorev.2014.03.015

Hardy, L., Mitchell, C., Seabrooke, T., and Hogarth, L. (2017). Drug cue reactivity involves hierarchical instrumental learning: evidence from a biconditional Pavlovian to instrumental transfer task. *Psychopharmacology (Berl.)* 234, 1977–1984. doi: 10.1007/s00213-017-4605-x

Hinterberger, T., Veit, R., Wilhelm, B., Weiskopf, N., Vatine, J.-J., and Birbaumer, N. (2005). Neuronal mechanisms underlying control of a brain-computer interface. *Eur. J. Neurosci.* 21, 3169–3181. doi: 10.1111/j.1460-9568.2005.04092.x

Hsueh, J.-J., Chen, T.-S., Chen, J.-J., and Shaw, F.-Z. (2016). Neurofeedback training of EEG alpha rhythm enhances episodic and working memory: alpha neurofeedback on Memory. *Hum. Brain Mapp.* 37, 2662–2675. doi: 10.1002/hbm.23201

Hu, L., Mouraux, A., Hu, Y., and Iannetti, G. D. (2010). A novel approach for enhancing the signal-to-noise ratio and detecting automatically event-related potentials (ERPs) in single trials. *NeuroImage* 50, 99–111. doi: 10.1016/j.neuroimage.2009.12.010

Ihssen, N., Sokunbi, M. O., Lawrence, A. D., Lawrence, N. S., and Linden, D. E. J. (2017). Neurofeedback of visual food cue reactivity: a potential avenue to alter incentive sensitization and craving. *Brain Imaging Behav.* 11, 915–924. doi: 10.1007/s11682-016-9558-x

Janapati, R., Dalal, V., Govardhan, N., and Gupta, R. S. (2020). Review on EEG-BCI classification techniques advancements. *IOP Conf. Ser. Mater. Sci. Eng.* 981:032019. doi: 10.1088/1757-899X/981/3/032019

Karelitz, J. L. (2020). Differences in magnitude of cue reactivity across durations of smoking history: a meta-analysis. *Nicotine Tob. Res.* 22, 1267–1276. doi: 10.1093/ntr/ntz071

Kirschner, M., Sladky, R., Haugg, A., Stämpfli, P., Jehli, E., Hodel, M., et al. (2018). Self-regulation of the dopaminergic reward circuit in cocaine users with mental imagery and neurofeedback. *EBioMedicine* 37, 489–498. doi: 10.1016/j.ebiom.2018.10.052

Koralek, A. C., Jin, X., Long, J. D. II, Costa, R. M., and Carmena, J. M. (2012). Corticostriatal plasticity is necessary for learning intentional neuroprosthetic skills. *Nature* 483, 331–335. doi: 10.1038/nature10845

Korostenskaja, M., Kapeller, C., Prueckl, R., Ortner, R., Chen, P. C., Guger, C., et al. (2014). "Improving ecog-based P300 speller accuracy," in *Proceedings of the 6th International Brain-Computer Interface Conference (Verlag der Technischen Universität Graz)*, 1–4. doi: 10.3217/978-3-85125-378-8-88

Kwon, J., Shin, J., and Im, C.-H. (2020). Toward a compact hybrid brain-computer interface (BCI): performance evaluation of multi-class hybrid EEG-fNIRS BCIs with limited number of channels. *PLoS One* 15:e0230491. doi: 10.1371/journal.pone.0230491

Li, X., Hartwell, K. J., Borckardt, J., Prisciandaro, J. J., Saladin, M. E., Morgan, P. S., et al. (2013). Volitional reduction of anterior cingulate cortex activity produces decreased cue craving in smoking cessation: a preliminary real-time fMRI study: Nicotine and real-time fMRI. *Addict Biol.* 18, 739–748. doi: 10.1111/j.1369-1600.2012.00449.x

Littel, M., Euser, A. S., Munafò, M. R., and Franken, I. H. A. (2012). Electrophysiological indices of biased cognitive processing of substance-related cues: a meta-analysis. *Neurosci. Biobehav. Rev.* 36, 1803–1816. doi: 10.1016/j.neubiorev.2012.05.001

Martz, M. E., Hart, T., Heitzeg, M. M., and Peltier, S. J. (2020). Neuromodulation of brain activation associated with addiction: a review of real-time fMRI neurofeedback studies. *NeuroImage Clin.* 27:102350. doi: 10.1016/j.nicl.2020.102350

Michel, C. M. (2018). EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: a review. *Neuroimage* 180(Pt. B), 577–593.

Posson, D. (2019). Automated neurofeedback brain-training as a primary addiction intervention. *Int. J. Educ. Soc. Sci.* 6, 21–27.

Rees, V. W., and Heather, N. (1995). "Individual differences and cue reactivity," in *Addictive Behaviour: Cue Exposure Theory and Practice. The Wiley Series in Clinical Psychology*, eds D. C. Drummond, S. T. Tiffany, S. Glautier, and B. Remington (Oxford: John Wiley & Sons), 99–118.

Sitaram, R., Ros, T., Stoeckel, L., Haller, S., Scharnowski, F., Lewis-Peacock, J., et al. (2017). Closed-loop brain training: the science of neurofeedback. *Nat. Rev. Neurosci.* 18, 86–100. doi: 10.1038/nrn.2016.164

Sokhadze, T. M., Cannon, R. L., and Trudeau, D. L. (2008). EEG Biofeedback as a Treatment for Substance Use Disorders: Review, Rating of Efficacy, and Recommendations for Further Research. *Appl. Psychophysiol. Biofeedback* 33, 1–28. doi: 10.1007/s10484-007-9047-5

Sokunbi, M. O. (2014). Real-time fMRI brain-computer interface: development of a "motivational feedback" subsystem for the regulation of visual cue reactivity. *Front. Behav. Neurosci.* 8:392.

Stam, C. J. (2010). Use of magnetoencephalography (MEG) to study functional brain networks in neurodegenerative disorders. *J. Neurol. Sci.* 289, 128–134. doi: 10.1016/j.jns.2009.08.028

Trambaiolli, L. R., Kohl, S. H., Linden, D. E. J., and Mehler, D. M. A. (2021). Neurofeedback training in major depressive disorder: a systematic review of clinical efficacy, study quality and reporting practices. *Neurosci. Biobehav. Rev.* 125, 33–56. doi: 10.1016/j.neubiorev.2021.02.015

Versace, F., Engelmann, J. M., Deweese, M. M., Robinson, J. D., Green, C. E., Lam, C. Y., et al. (2017). Beyond cue reactivity: non-drug-related motivationally relevant stimuli are necessary to understand reactivity to drug-related cues. *Nicotine Tob. Res.* 19, 663–669. doi: 10.1093/ntr/ntx002

Weber, E., Köberl, A., Frank, S., and Doppelmayr, M. (2011). Predicting successful learning of SMR neurofeedback in healthy participants: methodological considerations. *Appl. Psychophysiol. Biofeedback* 36, 37–45. doi: 10.1007/s10484-010-9142-x

Young, K. D., Misaki, M., Harmer, C. J., Victor, T., Zotev, V., Phillips, R., et al. (2017a). Real-time functional magnetic resonance imaging amygdala neurofeedback changes positive information processing in major depressive disorder. *Biol. Psychiatry* 82, 578–586. doi: 10.1016/j.biopsych.2017.03.013

Young, K. D., Siegle, G. J., Zotev, V., Phillips, R., Misaki, M., Yuan, H., et al. (2017b). Randomized clinical trial of real-time fMRI amygdala neurofeedback for major depressive disorder: effects on symptoms and autobiographical memory recall. *Am. J. Psychiatry* 174, 748–755. doi: 10.1176/appi.ajp.2017.16060637

Zafar, R., Dass, S. C., and Malik, A. S. (2017). Electroencephalogram-based decoding cognitive states using convolutional neural network and likelihood ratio based score fusion. *PLoS One* 12:e0178410. doi: 10.1371/journal.pone.0178410

Zhang, T., Chen, W., and Li, M. (2017). AR based quadratic feature extraction in the VMD domain for the automated seizure detection of EEG using random forest classifier. *Biomed. Signal Process. Control* 31, 550–559. doi: 10.1016/j.bspc.2016.10.001

Zhang, X. (2011). Anatomical differences and network characteristics underlying smoking cue reactivity. *Neuroimage* 54, 131–141.

Zhou, L., Wang, G., Nan, C., Wang, H., Liu, Z., and Bai, H. (2019). Abnormalities in P300 components in depression: an ERP-sLORETA study. *Nord. J. Psychiatry* 73, 1–8. doi: 10.1080/08039488.2018.1478991

Zotev, V., Phillips, R., Yuan, H., Misaki, M., and Bodurka, J. (2014). Self-regulation of human brain activity using simultaneous real-time fMRI and EEG neurofeedback. *NeuroImage* 85, 985–995. doi: 10.1016/j.neuroimage.2013.04.126

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership