

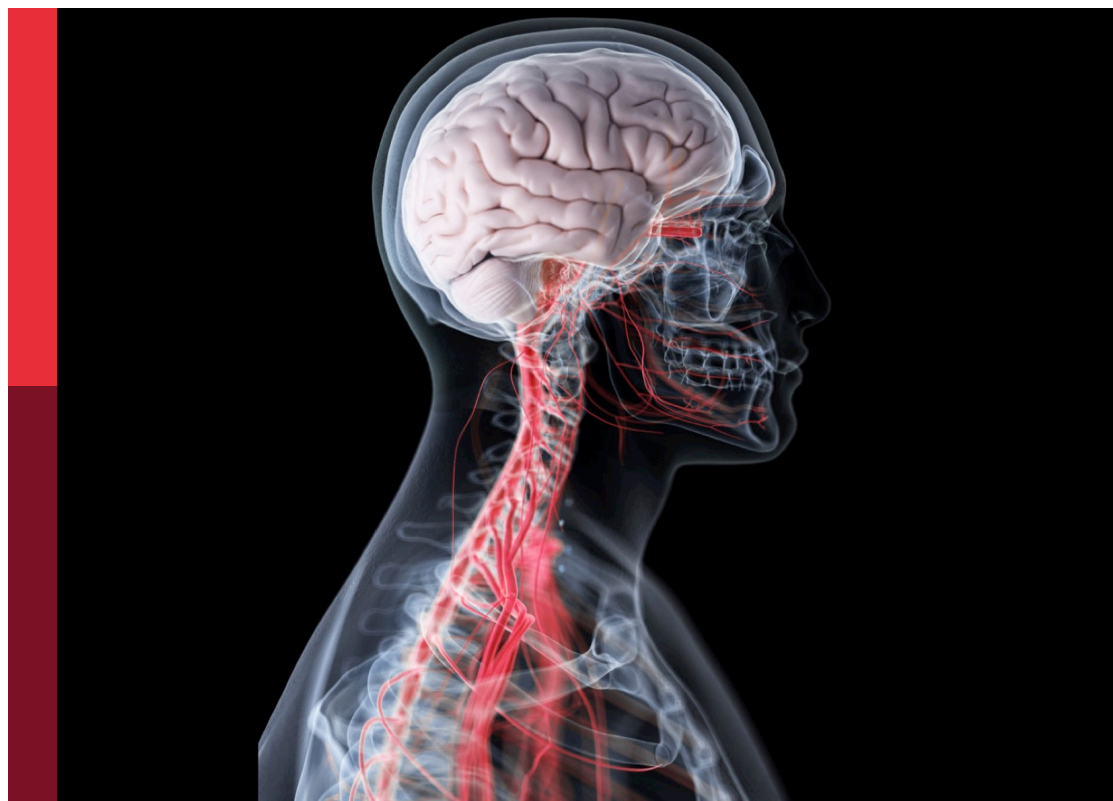
Translational brain-computer interfaces: From research labs to the market and back

Edited by

Davide Valeriani, Hubert Cecotti, Antonia Thelen and Christian Herff

Published in

Frontiers in Human Neuroscience



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-83251-764-2
DOI 10.3389/978-2-83251-764-2

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

Translational brain-computer interfaces: From research labs to the market and back

Topic editors

Davide Valeriani — Google, United States

Hubert Cecotti — California State University, Fresno, United States

Antonia Thelen — eemagine Medical Imaging Solutions GmbH, Germany

Christian Herff — Maastricht University, Netherlands

Citation

Valeriani, D., Cecotti, H., Thelen, A., Herff, C., eds. (2023). *Translational brain-computer interfaces: From research labs to the market and back*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83251-764-2

Table of contents

04	Editorial: Translational brain-computer interfaces: From research labs to the market and back Davide Valeriani, Hubert Cecotti, Antonia Thelen and Christian Herff
06	Like/Dislike Prediction for Sport Shoes With Electroencephalography: An Application of Neuromarketing Li Zeng, Mengsi Lin, Keyang Xiao, Jigan Wang and Hui Zhou
17	Target-Related Alpha Attenuation in a Brain-Computer Interface Rapid Serial Visual Presentation Calibration Daniel Klee, Tab Memmott, Niklas Smedemark-Margulies, Basak Celik, Deniz Erdogmus and Barry S. Oken on behalf of the Consortium for Accessible Multimodal Brain-Body Interfaces (CAMBI)
31	Brain-Heart Interaction and the Experience of Flow While Playing a Video Game Shiva Khoshnoud, Federico Alvarez Igarzábal and Marc Wittmann
48	Music, Math, and Working Memory: Magnetoencephalography Mapping of Brain Activation in Musicians Ching-I Lu, Margaret Greenwald, Yung-Yang Lin and Susan M. Bowyer
60	Usability of a Hybrid System Combining P300-Based Brain-Computer Interface and Commercial Assistive Technologies to Enhance Communication in People With Multiple Sclerosis Angela Riccio, Francesca Schettini, Valentina Galiotta, Enrico Giraldi, Maria Grazia Grasso, Febo Cincotti and Donatella Mattia
73	Evaluation of a New Lightweight EEG Technology for Translational Applications of Passive Brain-Computer Interfaces Nicolina Sciaraffa, Gianluca Di Flumeri, Daniele Germano, Andrea Giorgi, Antonio Di Florio, Gianluca Borghini, Alessia Vozzi, Vincenzo Ronca, Fabio Babiloni and Pietro Aricò
96	Frontal alpha asymmetry interaction with an experimental story EEG brain-computer interface Claudia Krogmeier, Brandon S. Coventry and Christos Mousas
114	Tangent space alignment: Transfer learning for Brain-Computer Interface Alexandre Bleuzé, Jérémie Mattout and Marco Congedo
126	The future of sensory substitution, addition, and expansion via haptic devices David M. Eagleman and Michael V. Perrotta



OPEN ACCESS

EDITED AND REVIEWED BY
Gernot R. Müller-Putz,
Graz University of Technology, Austria

*CORRESPONDENCE
Davide Valeriani
✉ davide.valeriani@gmail.com

SPECIALTY SECTION
This article was submitted to
Brain-Computer Interfaces,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 27 January 2023
ACCEPTED 31 January 2023
PUBLISHED 13 February 2023

CITATION
Valeriani D, Cecotti H, Thelen A and Herff C
(2023) Editorial: Translational brain-computer
interfaces: From research labs to the market
and back. *Front. Hum. Neurosci.* 17:1152466.
doi: 10.3389/fnhum.2023.1152466

COPYRIGHT
© 2023 Valeriani, Cecotti, Thelen and Herff.
This is an open-access article distributed under
the terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Editorial: Translational brain-computer interfaces: From research labs to the market and back

Davide Valeriani^{1*}, Hubert Cecotti², Antonia Thelen³ and
Christian Herff⁴

¹Google LLC, Mountain View, CA, United States, ²Department of Computer Science, California State
University, Fresno, CA, United States, ³eeimagine Medical Imaging Solutions GmbH, Berlin, Germany,
⁴Department of Neurosurgery, Maastricht University, Maastricht, Netherlands

KEYWORDS

brain-computer interface, neurotechnology and brain-machine interface, EEG, fNIRS,
neuromarketing, BCI, human-computer interaction

Editorial on the Research Topic

Translational brain-computer interfaces: From research labs to the market and back

Neurotechnologies combine neuroscience and engineering to build systems for studying, repairing, and augmenting human performance. These include brain-computer interfaces (BCIs)—typically used as assistive devices for communication and rehabilitation for disabled people and, more recently, as human enhancement devices—neural prosthetics, optogenetics, digital medicine, continuous remote monitoring, drugs, and other pharmaceutical interventions.

This Research Topic provides a collection of nine novel contributions on recent bidirectional interactions between academic labs and industry partners to showcase translational work in BCIs for clinical, business-to-business, and consumer markets that push neurotechnologies outside of the lab.

In the context of neuromarketing, Zeng et al. showed that whether users like or dislike a pair of sport shoes could be understood by looking at their electroencephalography (EEG) activity. Amongst all feature extraction methods explored, they found that differential entropy features extracted from signals recorded from the occipital region provided the highest classification accuracy in detecting shoes that the user liked. The importance of the occipital cortex in visual processing and emotion arousal is well-known in neuroscience literature (Lang et al., 1998). An auditory task may instead have involved the temporal region more heavily than the occipital one. The authors demonstrated that the frontal region also played an important role in boosting classification accuracy, as the decision-making process of the user occurs in this area (Volz et al., 2006). This study makes a first step toward using passive BCI systems as tools for gathering user feedback on consumer products, without specifically asking to rate them through traditional reviews. Future studies need to explore if similar results could be obtained using consumer EEG devices with dry electrodes, instead of research grade wet EEG systems as used in the study.

The process of porting BCIs from research labs to the market encompasses the development of easy to use, reliable, and cost effective sensors that could replace the usually bulky, expensive, and difficult to setup research-grade EEG systems. An attempt to respond to this need has been made by [Sciaraffa et al.](#), who have designed and tested a BCI system composed of water-based electrodes and a lightweight headset. They found this system to provide comparable results to BCIs based on research-grade EEG systems and gel-based electrodes across a variety of passive tasks and environments, including multitasking, psychomotor vigilance task, and car driving simulation. The usage of water electrodes increased ease of use without lowering performance, and the lightweight headset made the BCI comfortable enough for many real-life applications. Another great challenge in BCI translation is reducing the training time, i.e., the calibration session, enabling users to benefit from the BCI as soon as possible. A key technique in this domain is transfer learning, namely reusing some of the information gathered from a previous training session to train a new model for a different task or user. [Bleuzé et al.](#) exploited Riemannian geometry to develop a new transfer learning method for BCIs that identifies who is the most similar subject to the user at hand in the previous training session, and uses these parameters for the new model. They extensively tested this approach on 18 BCI databases, finding a significant improvement on BCI performance over standard training-test pipelines, including those using Riemannian geometry.

A typical BCI application out of the labs is their use in entertainment. A study in this Research Topic explored the interaction between brain and heart through the heart-evoked potential in gamers and non-gamers playing a video game ([Khoshnoud et al.](#)). They found that flow—the state of complete absorption in an activity—can be tracked through EEG sensors and does not differ between gamers and non-gamers, as long as the user is able to set the level of difficulty of the video game. Another study tracked the process of musical transposing (i.e., converting notes across musical keys) in musicians and found it significantly differs from math calculations, although both tasks involve the usage of working memory ([Lu et al.](#)). Another novel study explored the usage of BCIs for story generation through increases of frontal alpha asymmetry ([Krogmeier et al.](#)). Although only 37% of the participants were able to successfully modulate their brain activity during story creation, this novel paradigm further reveals the importance of understanding BCI literacy and what factors make users able to control a novel BCI successfully.

As BCIs were born as human-computer interaction technologies, another great area of impact of BCIs out of the lab is helping patients with serious neurological disorders regain their independence. A pilot study by [Riccio et al.](#) explored the use of BCIs in patients with multiple sclerosis, as a possible integration with other assistive devices designed for these patients.

The majority of the patients were able to control an assistive device software with their brain activity, paving the way to the integration of BCIs into the daily life of multiple sclerosis patients. While [Riccio et al.](#) used an established paradigm of BCI based on the detection of event-related potentials (the P300 speller), [Klee et al.](#) explored the usage of occipitoparietal alpha activity in a speller based on rapid serial visual presentation (RSVP). This novel paradigm for communication performed better than random, but well-below more established paradigms for spellers. Future investigations should test the possibility of combining these paradigms into a hybrid approach, which may lead to superior performance than other paradigms. Combining different types of assistive devices seem to be the next frontier for BCIs, as presented by [Eagleman and Perrotta](#) in their perspective on the future of sensory substitution, addition and expansion *via* haptic devices, which indeed represent very promising technologies for human rehabilitation and augmentation.

Author contributions

DV drafted the manuscript. All authors provided feedback and contributed to the Research Topic. All authors contributed to the article and approved the submitted version.

Acknowledgments

We hope this Research Topic provides the reader with updates on recent advances in translational BCIs. We would like to thank all authors who contributed, and the reviewers who provided invaluable and timely feedback to the authors.

Conflict of interest

DV is an employee at Google LLC. AT is employed by eemagine GmbH.

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.

References

Lang, P. J., Bradley, M. M., Fitzsimmons, J. R., Cuthbert, B. N., Scott, J. D., Moulder, B., et al. (1998). Emotional arousal and activation of the visual cortex: an fMRI analysis. *Psychophysiology* 35, 199–210.

Volz, K. G., Schubotz, R. I., and von Cramon, D. Y. (2006). Decision-making and the frontal lobes. *Curr. Opin. Neurol.* 19, 401–406. doi: 10.1097/01.wco.0000236621.83872.71



Like/Dislike Prediction for Sport Shoes With Electroencephalography: An Application of Neuromarketing

Li Zeng^{1,2}, Mengsi Lin³, Keyang Xiao³, Jigan Wang¹ and Hui Zhou^{3*}

¹ School of Business, Hohai University, Nanjing, China, ² College of Environment, Hohai University, Nanjing, China, ³ School of Automation, Nanjing University of Science and Technology, Nanjing, China

OPEN ACCESS

Edited by:

Vilfredo De Pascalis,
Sapienza University of Rome, Italy

Reviewed by:

Ameersing Luximon,
EMEDS, Hong Kong SAR, China
Joseph Ciorciari,
Swinburne University of Technology,
Australia

*Correspondence:

Hui Zhou
zhouhui@njust.edu.cn

Specialty section:

This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

Received: 12 October 2021

Accepted: 26 November 2021

Published: 06 January 2022

Citation:

Zeng L, Lin M, Xiao K, Wang J
and Zhou H (2022) Like/Dislike
Prediction for Sport Shoes With
Electroencephalography: An
Application of Neuromarketing.
Front. Hum. Neurosci. 15:793952.
doi: 10.3389/fnhum.2021.793952

Neuromarketing is an emerging research field for prospective businesses on consumer's preference. Consumer's preference prediction based on electroencephalography (EEG) can reliably predict likes or dislikes of a product. However, the current EEG prediction and classification accuracy have yet to reach ideal level. In addition, it is still unclear how different brain region information and different features such as power spectral density, brain asymmetry, differential entropy, and Hjorth parameters affect the prediction accuracy. Our study shows that by taking footwear products as an example, the recognition accuracy of product likes or dislikes reaches 94.22%. Compared with other brain regions, the features of the frontal and occipital brain region obtained a higher prediction accuracy, but the fusion of the features of the whole brain region could improve the prediction accuracy of likes or dislikes even further. Future work would be done to correlate the EEG-based like or dislike prediction results with product sales and self-reports.

Keywords: neuromarketing, electroencephalography, machine learning, brain asymmetry, preference prediction

INTRODUCTION

Neuromarketing is an emerging interdisciplinary research area that aims to understand biology of consumer's behavior by integrating neuroscience with marketing, which can decipher consumers' unrevealed preferences, motivations, and decisions by measuring their physiological and neural signals (Ariely and Berns, 2010; Morin, 2011; Aldayel et al., 2020; Bazzani et al., 2020). It is estimated that neuromarketing has market potential of 400 billion dollars (Khurana et al., 2021). Conventional marketing provides only relative analysis of consumer's response, which relies on conducting surveys, interviews, running focus groups, and field trials for collecting consumer's feedback. These analysis approaches suffer limitations due to high cost, time requirement, and untrustworthy information. Besides, conventional approaches have significant inherent weaknesses arising from consumers not always forthcoming about their feelings and preferences. All of these drawbacks would lead to biased or inaccurate conclusions (Khushaba et al., 2012; Boksem and Smidts, 2015). Compared with conventional marketing research techniques, neuromarketing empowers researchers to capture consumers' intricate neural processes to a range of marketing stimuli with moment-to-moment neural data, allowing to forecast consumer's decision-making, like-dislike, and purchase decisions with greatly improved accuracy (Venkatraman et al., 2012; Barnett and Cerf, 2017; Bell et al., 2018; Goto et al., 2019).

In neuromarketing, neural signal recording techniques are commonly used to directly measure consumer's brain responses to the marketing stimuli (Ariely and Berns, 2010; Sánchez-Fernández et al., 2021). Popular noninvasive neuroscientific techniques to analyze and understand consumer's behavior include brain imaging technologies such as electroencephalography (EEG), magneto-encephalography (MEG), functional magnetic resonance imaging (fMRI), functional near-infrared spectroscopy (fNIRS), and various physiological parameters (e.g., heart rate and respiratory rate) (Khurana et al., 2021; Qing et al., 2021). For neuromarketing, EEG has several benefits that other physiological signals lack, such as high temporal resolution for detecting brain activity changes at low cost, equipment portability. Because it can be easily employed in real-time marketing environments, it is desirable to use EEG to capture electrical brain activity and assess marketing stimuli to build preference prediction system in the neuromarketing research (Aldayel et al., 2020).

For companies and advertisers, it is of great significance to successfully predict consumer's preference of specific products. They can reduce inventory, increase profits, grow customer loyalty, and satisfaction and improve branding by competently avoiding the production of unpopular or undesirable products. Consumer's preferences of footwear products were proficiently predicted using EEG data as compared to self-report-based predictions (Baldo et al., 2015). Yilmaz et al. (2014) researched consumers' likes and dislikes of footwear products using EEG signals and revealed the particularly different channels and frequencies of likes and dislikes. In neuromarketing, power spectral density (PSD) is one of the most common feature extraction method (Khushaba et al., 2012; Yilmaz et al., 2014). Some studies theorize that the PSD of EEG signals can be used to identify the likes and dislikes of products (Golnar-Nik et al., 2019). Additionally, some studies presume that features and parameters of frontal brain asymmetry, such as approach-withdrawal (AW) index, effort index, choice index, valence, can expertly identify product preference of consumers (Cartocci et al., 2017; Ramsøy et al., 2018; Aldayel et al., 2020, 2021). Differential entropy (DE) and Hjorth parameters are commonly used EEG features in emotion recognition (Chen et al., 2019; Joshi and Ghongade, 2022). However, as far as we know, the use of these features in neuromarketing has seldom been reported.

It is often challenging for neuromarketing researchers to choose the appropriate characteristics from these many EEG characteristics to accurately predict product likes or dislikes (Bell et al., 2018). Therefore, it is necessary to carry out the research to compare the classification accuracy of these features, so that further studies in the field of neuromarketing can be done to apply to more product like and dislike cases.

The purpose of this study is to design a product like or dislike prediction system based on EEG using commonly worn sport shoes as product example, so that comparison of the characteristics of consumer's preference of commonly used EEG in published literature can be made and also to compare the classification accuracy of these features. This can help

neuromarketing scholars to design a classification and prediction system based on EEG. The contributions of this research are as follows:

- (1) Developing a consumer's like or dislike prediction system based on EEG to achieve high classification accuracy by taking sport shoes as an example.
- (2) Study the influence of different EEG characteristic parameters such as different brain locations, PSD features, brain asymmetry features, Hjorth features, and DE features and compare their classification accuracies.

Based on the abovementioned criteria, our study looks to implement method and provide test results of the EEG-based shoe like or dislike prediction system in the following chapters. To attain our goal, EEG-based preference detection experiments were conducted. During the experiment, electrical brain activity of 15 subjects aged between 22 and 39 was recorded. Subjects were presented with pictures of 25 different sport shoes one by one and were asked to decide whether they liked or not by pressing 1 on keyboard for like and 2 on keyboard for dislike. In session II, the details of participants, trial design, experimental equipment, and machine learning classifier framework are explained. In session III, the PSD, brain asymmetry, Hjorth parameter, and DE were extracted as features. Machine learning classifiers SVM and KNN are used, and finally, the system is evaluated by measuring the accuracies of the classifier.

MATERIALS AND METHODS

Subjects

Fifteen healthy subjects (nine men and six women, 22–39 years old, all right-handed) who were students of Nanjing University of Science and Technology participated in the experiment. All participants reported normal hearing and the absence of any neurological disorders. They were informed about the purpose and experimental procedure of the study. Both genders were participated in the experiment to explore possible influence of gender in results. The recruitment of subjects and the experimental protocols were approved by the Ethics Committee for Human Research, Nanjing Brain Hospital Affiliated to Nanjing Medical University.

Experiment Protocol

A computerized task has been designed to investigate the ability of EEG power to distinguish between consumer's preferences among subjects. Participants were asked to make a decision for liking or disliking of specific product. According to 2020 online shopping report of China, the most favorable commodities among online consumers are daily necessities, clothing and shoes. Therefore, for this study, pictures of 25 different sport shoes under the Chinese Li-Ning brand were selected in the experiment as shown in **Figure 1**.

The experimental paradigm is implemented using E-Prime software. **Figure 2** displays the entire process of experimental protocol. The main part of the experiment is divided into 25 runs (iterations/epochs). Each run lasts for 11 s, and it includes



FIGURE 1 | Images of 25 different sport shoes used in the experiment.

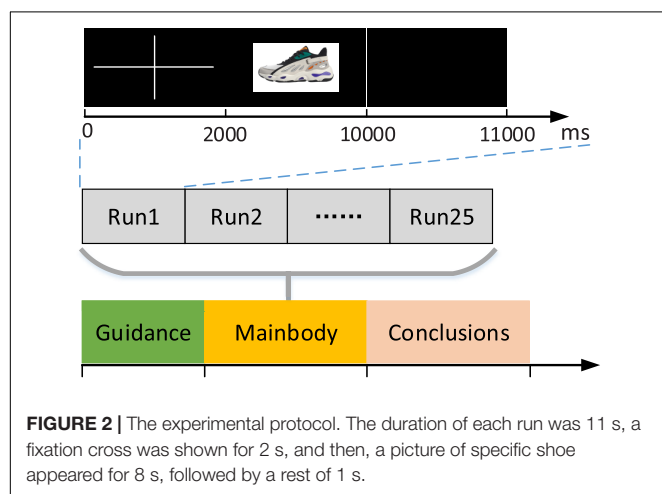
prompting to focus attention for 2 s, observing image of specific shoe for 8 s, and resting for 1 s. Each round will randomly present a shoe image without repetition, and the subjects were asked to look at the product image. In the displayed 8 s, subjects make a decision on whether they like the shoe or not and record

their decision through the keyboard by pressing “1” for like and “2” for dislike.

Data Acquisition and Data Preprocessing

During the experiment, a 32-channel EEG acquisition device, BrainAmp Amplifier (Brain Product, Gilching, Germany) and active Ag/AgCl electrodes (actiCAP, Brain Product, Germany), was used to collect the EEG signals of the subjects. For each run (epoch), the EEG data are recorded from 22 silver chloride electrodes placed on subject's head in the international 10–20 system. The specific placement locations are as follows: FP1, FPz, FP2, F7, F3, Fz, F4, F8, T7, C3, Cz, C4, T8, TP9, TP10, P7, P3, Pz, P4, P8, O1, O2, with FPz as the ground electrode and Fz as the reference electrode. The conductive paste is used to make the contact impedance between the electrode and the scalp less than 5 k Ω , the sampling rate is set to 500 Hz, and the bandpass filter was set at 0.03–70 Hz.

To accurately collect effective EEG data, the participants were asked to ensure adequate sleep before the experiment, stimulants such as cigarettes, alcohol, coffee, and strenuous exercise should be avoided prior to testing, and hair should be cleaned. The environment was kept quiet during the experiment. The participants sat in a comfortable chair, looked at the screen



squarely, kept their mind and body in relaxed state, maintained their posture, and reduced the number of eye blinks.

Brain analyzer is used to preprocess raw EEG data recorded in the experiment, including re-referencing, filtering, removing artifacts, segmenting, and reducing the sampling rate. First, we re-referenced the data. The current electrode caps usually use Fz as the reference electrode, but in the analysis of EEG data, the reference electrode needs to be replaced according to the experimental requirements. The reference electrodes used in this analysis are bilateral mastoid TP9 and TP10. Then, we filtered the data. The raw EEG signals were bandpass filtered using a 4th-order Butterworth filter set at 0.5–40 Hz to filter out high-frequency noise. We need to carry out the necessary artifact removal operation for EEG signal. As the artifact signals caused by the device or the subject's actions will cause errors in subsequent data processing and experimental results, this article uses independent component analysis (ICA) to correct physiological artifacts such as electrooculogram (EOG). Then, we need to segment the data into likes and dislikes. According to the markers formed on the EEG data of the subjects' choice reaction to the product during the experiment, the data of the state of like and dislike are extracted separately. Finally, the data are sampled down. To reduce the amount of data and increase the calculation speed, the sampling rate is reduced to 256 Hz.

Feature Extraction

To select appropriate features, different feature extraction technologies are used and compared for predicting like and dislike of a shoe product. The EEG features of PSD, brain asymmetry, DE, and Hjorth parameter were chosen in this study. In neuromarketing, PSD is one of the most common feature extraction methods (Khushaba et al., 2012; Yilmaz et al., 2014; Golnar-Nik et al., 2019). The brain asymmetry-based preference indices such as approach-withdrawal (AW) index, valence, choice index, and effort index are also used as features to predict consumer's preference (Aldayel et al., 2020, 2021). Besides, EEG features such as DE and Hjorth parameters have generally been used in some EEG-based applications (Chen et al., 2019; Joshi and Ghongade, 2022). However, as far as we know, the use of these features in neuromarketing has seldom been reported.

Power Spectral Density

The PSD is an indicator of power in a certain signal in terms of frequency. The Welch method is used to estimate the PSD of the EEG signal, and the PSD of the time series is calculated as below (Welch, 1967):

$$\hat{S}_{XX}(k) = \frac{1}{N} \sum_{n=1}^N |X_n(k)|^2 \quad (1)$$

where $X_n(k)$ is the Fourier transform of the time series x corresponding to the n th segment and the k th frequency point after windowing. In this study, the relative power in the four frequency bands of δ (0.5–4 Hz), θ (4–8 Hz), α (8–13 Hz), and β (13–30 Hz) is calculated based on the PSD of each channel data.

Brain Asymmetry

The AW index of frontal alpha asymmetry estimates desire and motivation as alpha's higher activation in the left frontal cortex. We can measure the AW scores using electrodes F4 and F3 to find the difference between the right and left PSD divided by their amounts according to Eq. 2 (Touchette and Lee, 2017).

$$AW \text{ index} = \frac{\alpha(F4) - \alpha(F3)}{\alpha(F4) + \alpha(F3)} \quad (2)$$

The effort index measures effort and cognitive processing as higher theta activation in the prefrontal cortex. We used the following equation to calculate the effort index (Aldayel et al., 2021).

$$\text{Effort index} = \frac{\theta(F4) - \theta(F3)}{\theta(F4) + \theta(F3)} \quad (3)$$

The choice index is defined in Eq. 4. The choice index can be calculated for each band individually using electrodes pairs of left and right counterparts for each lobe according to Eq. 4 (Moon, 2013; Ramsøy et al., 2018).

$$\text{Choice index} = \frac{\log(\text{Electrode}_{\text{left}}) - \log(\text{Electrode}_{\text{right}})}{\log(\text{Electrode}_{\text{left}}) + \log(\text{Electrode}_{\text{right}})} \quad (4)$$

The valence measures positive emotion as left frontal activation in alpha and beta bands. In this study, we computed the values of valence using Eq. 5 (Al-Nafjan et al., 2017).

$$\text{Valence} = \frac{\alpha(F4)}{\beta(F4)} - \frac{\alpha(F3)}{\beta(F3)} \quad (5)$$

Differential Entropy

Differential entropy is defined in Eq. 6, where $p(x)$ represents the probability density function of continuous information (Shi et al., 2013).

$$DE = - \int_a^b p(x) \log(p(x)) dx \quad (6)$$

For EEG signal with a specific length that approximately follows a Gaussian distribution, its DE is expressed as follows:

$$DE = - \int_a^b \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x-\mu)^2}{2\sigma_i^2}} \log\left(\frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x-\mu)^2}{2\sigma_i^2}}\right) dx \quad (7)$$

$$= \frac{1}{2} \log(2\pi e \sigma_i^2)$$

Hjorth Parameters

Hjorth introduced the Hjorth parameters to describe the EEG signal in the time domain, including the following three characteristics, which are activity, mobility, and complexity (Hjorth, 1970):

Activity measures the degree of deviation of the signal amplitude:

$$\text{Activity} = \frac{1}{N} \sum_{n=1}^N (s(n) - \mu_s)^2 \quad (8)$$

Mobility measures the changes in slope:

$$\text{Mobility} = \sqrt{\frac{\text{var}(s'(n))}{\text{var}(s(n))}} \quad (9)$$

Complexity measures how many standard slopes are there on an amplitude:

$$\text{Complexity} = \frac{\text{Mobility}(\dot{s}(n))}{\text{Mobility}(s(n))} \quad (10)$$

where μ_s represents the average value of the signal, $\dot{s}(n)$ represents the first derivative of the signal, and $\text{var}(\cdot)$ represents the variance.

Classification and Statistical Analysis

This study implements k-nearest neighbor (KNN) and support vector machine (SVM) classifiers to distinguish the EEG characteristics of consumer's preference between likes and dislikes based on KNN with Euclidean distance [KNN(E)], KNN with cosine distance [KNN(C)], SVM with radial basis kernel [SVM(R)], and SVM with polynomial kernel [SVM(P)] since these are commonly used machine learning methods. During the experiment, the EEG signals are collected when the subjects watch pictures different sport shoes. The 10-fold crossvalidation was performed, and all subjects' data were collected and randomly split into training set (90%) and test set (10%). At the same time, the data of two data sets are preprocessed and extracted. Among the features extracted were PSD, brain asymmetry, DE, and Hjorth parameters. Then, the parameters of the classifier model are trained by the feature data and labels in the training set, and then, the performance of the

trained model is evaluated by the feature data and labels in the test set, as shown in **Figure 3**.

The classification accuracy was defined as:

$$\text{Accuracy} = \left(\frac{N_{\text{correct}}}{N_{\text{total}}} \right) \times 100\% \quad (11)$$

where N_{total} and N_{correct} are defined as the total number of samples to be classified and the number of correct samples. The final classification accuracy was the average of 10 repetitions.

For statistical analysis of difference in power of EEG between like and dislike decision, two-sample *t*-tests are performed by calculating the power of different frequency bands. The *p*-value (≤ 0.05) from two-sample *t*-tests represent the significant contrast between liked and disliked decisions.

RESULTS

Most Liked or Disliked Sport Shoes

Figure 4 displays the most liked and disliked shoes among the following subject groups: male subjects only, female subjects only, and all subjects. **Figure 4A** displays most liked shoes among male subjects. The shoe farthest on the left is liked by seven male subjects, and the other three are equally liked by six male subjects. Almost all male subjects disliked the shoe styles in **Figure 4B**. The two shoe styles in **Figure 4C** were equally liked by five female subjects, respectively, and all of the female subjects disliked the 8

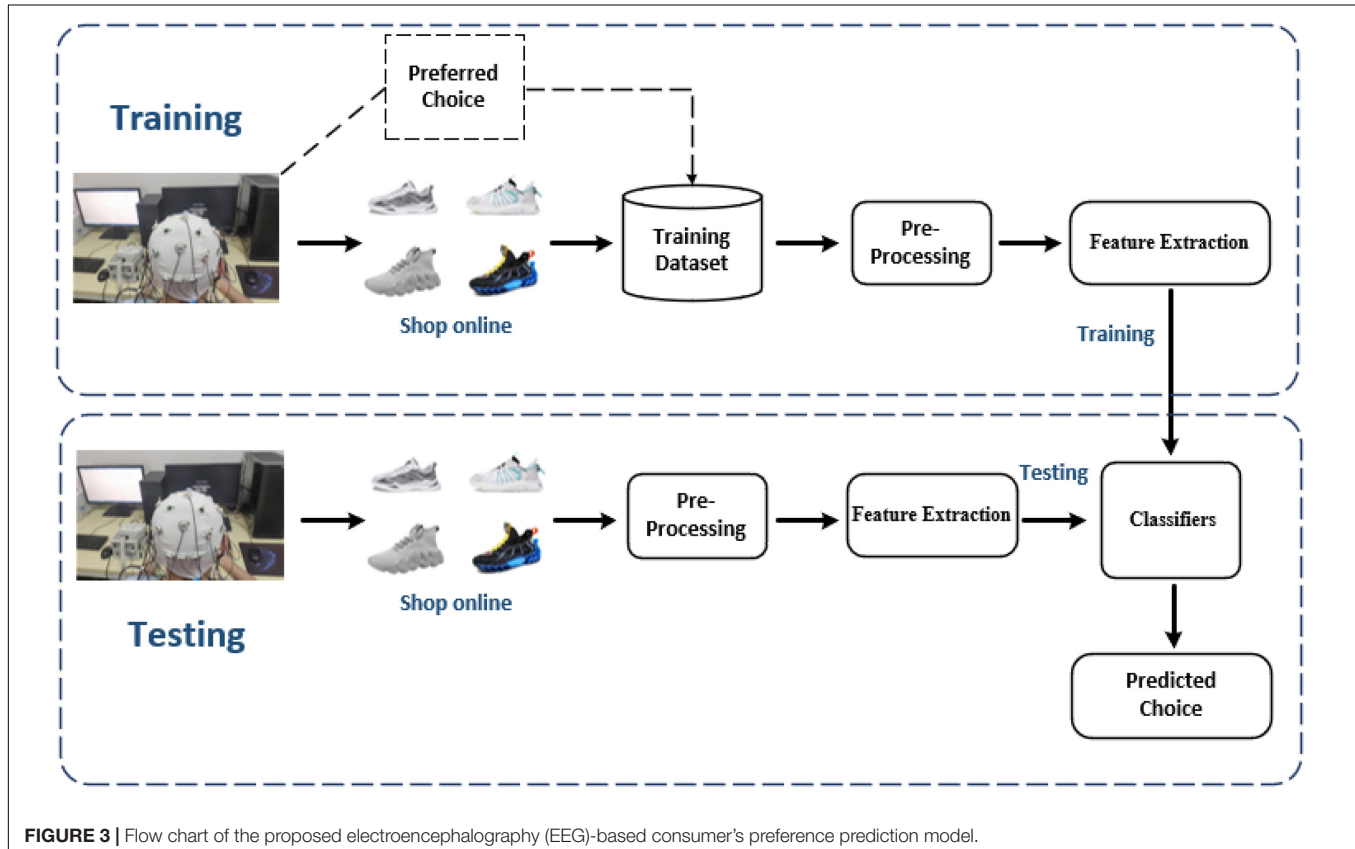




FIGURE 4 | Most liked and disliked shoes among different subject types (men, women, and all). Panel (A) are the most liked shoes among male subjects, panel (B) are the most disliked shoes among male subjects, panel (C) shows most liked shoes among female subjects, panel (D) shows most disliked shoes among female subjects, panel (E) are overall most liked shoes among all subjects, and panel (F) are overall most disliked shoes among all subjects.

shoe styles in **Figure 4D**. Finally, we obtained the two shoes which all subjects (men and women combined) like and dislike the most. In **Figure 4E**, the shoes that were liked by 12 and 11 subjects are shown from left to right. In **Figure 4F**, the two most disliked shoes are shown, each of which was liked by only one subject.

Power Spectral Density Analysis

The PSD of 0.5–40 Hz was calculated for 2,250 samples (25 runs \times 15 subjects \times 6 segments of 8-s data) of all subjects. The results were averaged according to the label. **Figure 5** shows the PSD result where the upper and lower parts represent the like and dislike.

Figure 6 represents the significant power difference between like and dislike choices made by subjects. In the delta frequency band, significant EEG power difference between like and dislike can be observed in the right frontal, left temporal, right temporal, and right occipital regions. For the theta frequency band, the significant power difference is concentrated in temporal, parietal, central, and frontal in asymmetry manner. Significant power difference in the alpha frequency band is concentrated in the frontal (symmetry) and parietal areas. In addition, the significant power difference in the beta frequency band is found in the frontal, temporal, central, parietal, and occipital regions. These findings are similar to other studies (Golnar-Nik et al., 2019).

Classification Result

Figure 7 shows the classification results of four different classifiers based on four different feature sets. It can be observed from the classification accuracy that the two classifiers KNN(C) and KNN(E) give better performance than the two classifiers SVM(P) and SVM(R) based on PSD features, choice-based symmetry features, Hjorth parameter features, and the DE features. KNN(C) classifier gives the best results where the accuracy is 88.85% for PSD features, 82.04% for choice-based

asymmetry features, 86.17% for Hjorth features, and 94.22% for DE features. Note that 16 channels were used for choice-based asymmetry features.

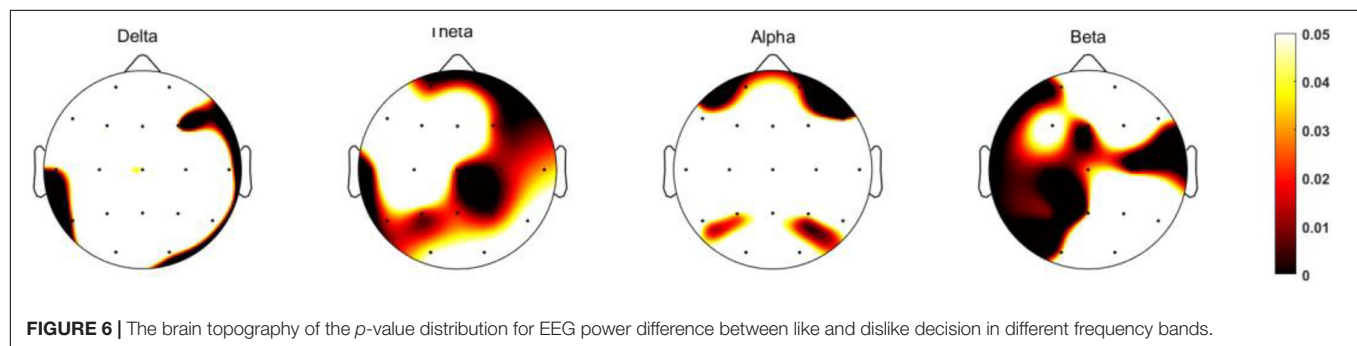
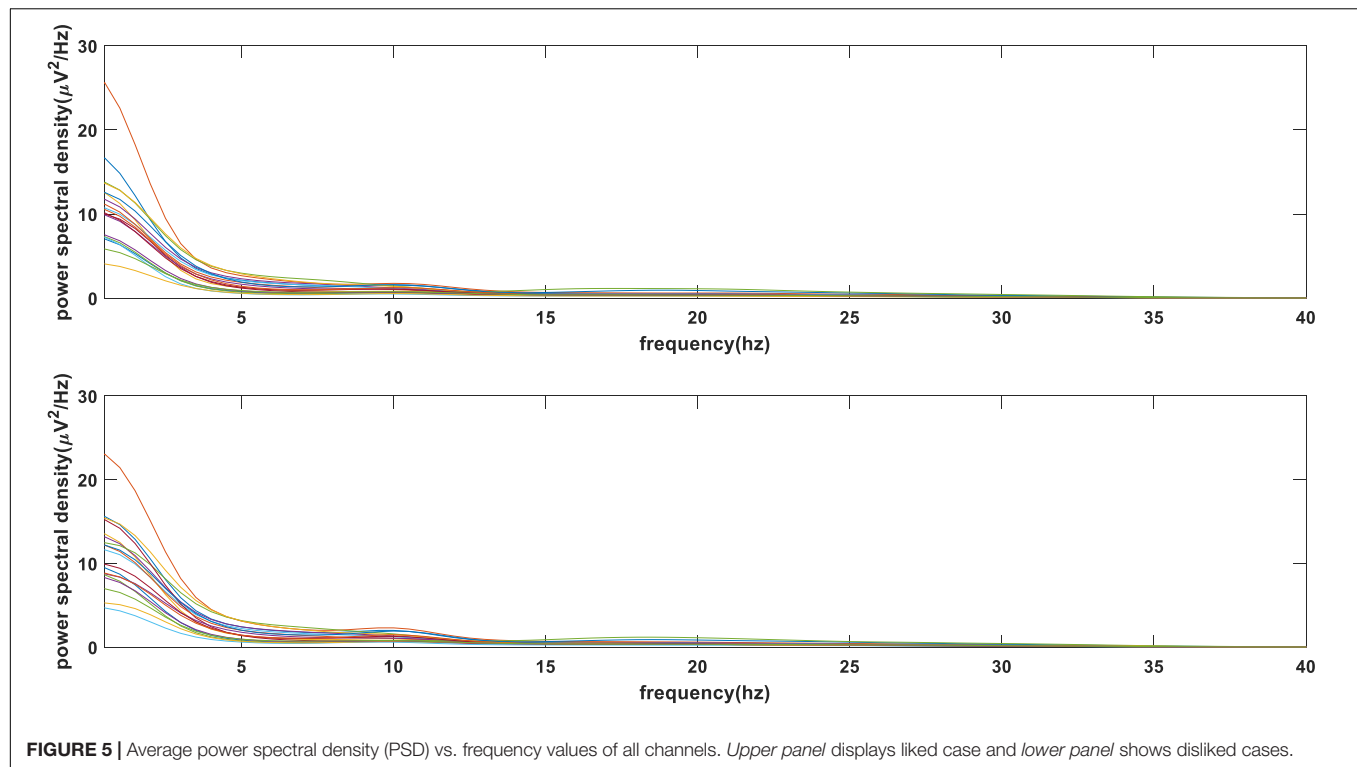
Figure 8 shows the classification results of different brain lobes based on different feature sets, and the experimental results shown below are based on the KNN(C) classifier. The results showed that the difference in like and dislike of shoe products was most apparent in the DE features of the occipital locations (87.16%). However, fusion of all brain region areas increases classification accuracy of all four feature sets.

Figure 9 shows the classification results of different frequency bands based on different feature sets, and the experimental results shown below are based on the KNN(C) classifier. Compared with other frequency bands, the results show that alpha bands of signals are slightly more conducive to the distinction between consumer's likes and dislikes. Besides, the features of other frequency bands can also be used to achieve like and dislike classification. The fusion of all frequency bands resulted in improved classification accuracies of like and dislike prediction.

Figure 10 shows comparison of the classification accuracies of features of different EEG indices based on the KNN(C) classifier. It is observable that all four EEG asymmetry-based features (AW index, effort index, choice index, and valence) provide similar classification accuracy when the same channels were used (F4 and F3). However, 16 channel choice index-based features give higher accuracy than 2 channel choice features and rest of the EEG index-based features. Nevertheless, features of PSD (88.85%), Hjorth (86.17%), and DE (94.22%) provide highest classification accuracy and noticeably greater than the rest.

DISCUSSION

In this research, an EEG-based consumer's preference prediction system is proposed to predict whether consumers' like or dislike



particular product. EEG data were recorded from 15 recruited subjects whereas they watched pictures of 25 different designs of sport shoes and made decisions about their likes and dislikes of shoes that were shown. The EEG-based consumer's like or dislike prediction system demonstrated in this work achieved a classification accuracy of 94.22% using DE features.

Our research weighs up the effects of the characteristics of different brain regions on classification accuracy. It can be observed from the p -value distribution of EEG power that the frontal and occipital regions play important part in consumer's like and dislike decision-making process. Furthermore, the features extracted from those two regions also may have higher classification accuracy than the other brain regions.

Electroencephalography-based preference indices such as AW index, effort index, choice index, and valence are often used to measure the response of subjects to market stimuli. However, the classification accuracy obtained from the features extracted from these indices is seldom reported. Our research finds that

classification accuracy of these indices is similar to accuracy of choice index being slightly higher. Increasing the number of channels sees increase in classification accuracy among all EEG features, as evidenced by the difference between 2 channel and 16 channel choice index features. Our research implemented a method that demonstrated extracted DE features which were able to obtain greatly higher classification accuracy (>90%) for both KNN(C) and KNN(E) classifiers. Besides, combining features of all frequency bands see improvements in classification accuracy.

However, there are some limitations in this study. First, only product images were used as marketing stimuli, and other factors such as brand, ratings, and price were not considered in the experiment. Second, there were only 15 subjects of college students recruited in the experiment for consumer's preference prediction of sport shoes. Increasing the number of subjects of different ages, incomes, and social status would help to investigate the influence of different

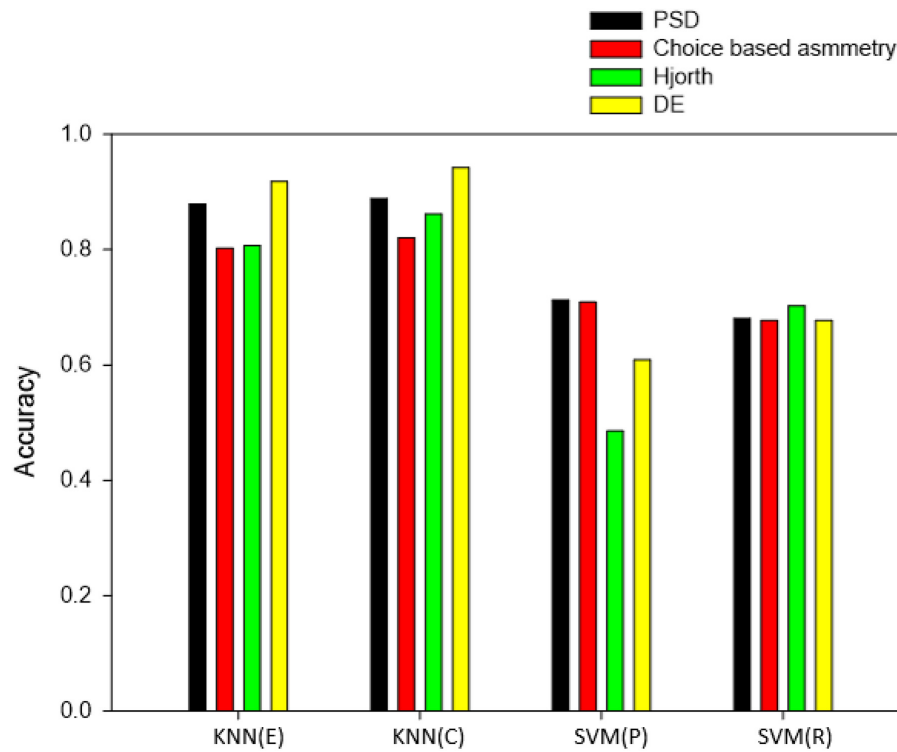


FIGURE 7 | Classification accuracy of four classifiers based on different feature sets.

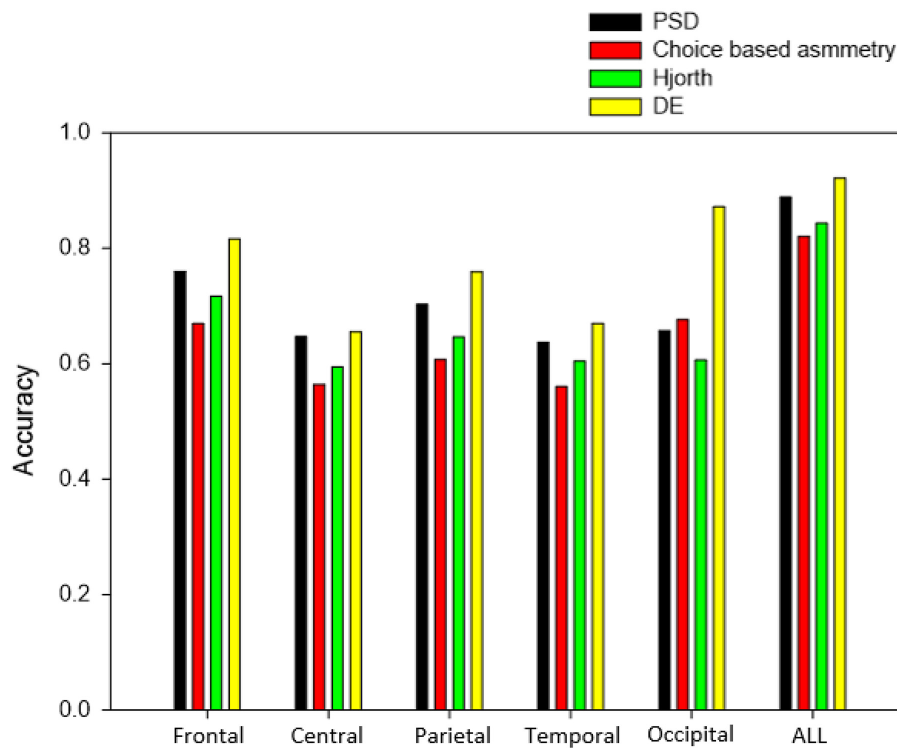


FIGURE 8 | Classification accuracy of different brain lobes based on different feature sets.

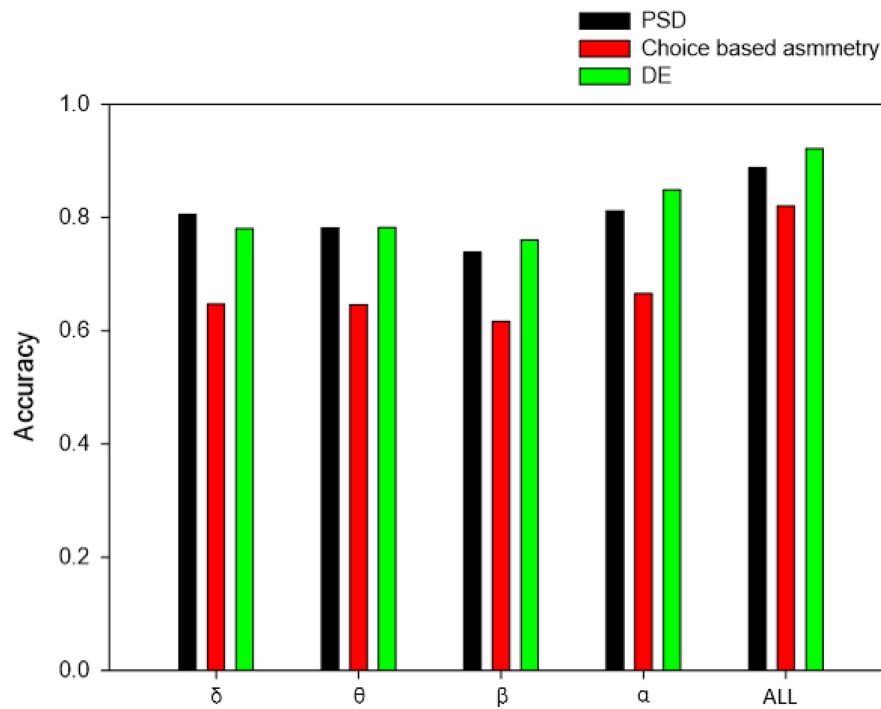


FIGURE 9 | Classification accuracy of different frequency bands based on different feature sets.

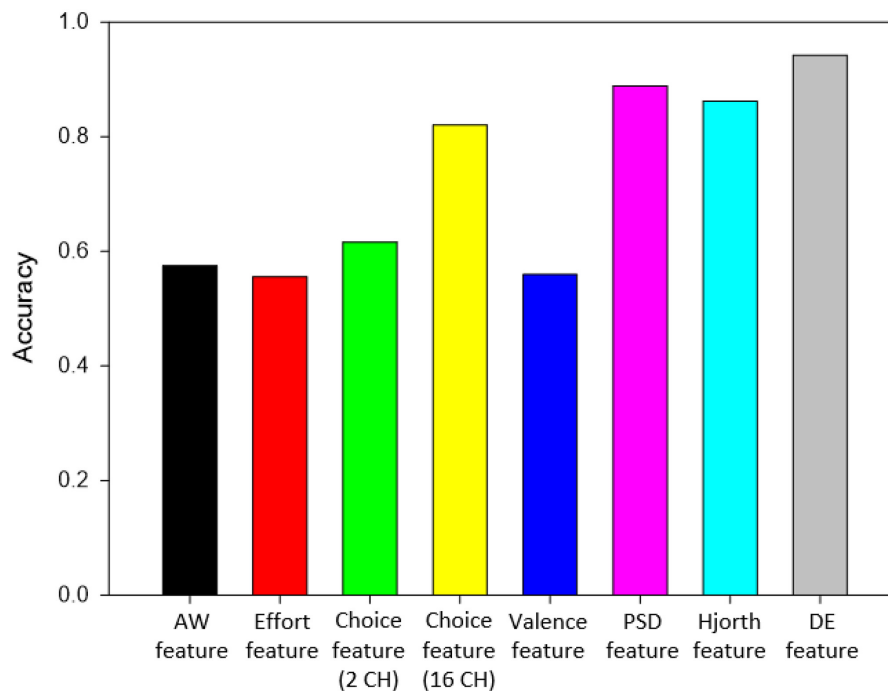


FIGURE 10 | Classification accuracy of different EEG features.

factors (brand and ratings, etc.) on consumer's preferences (Golnar-Nik et al., 2019). Furthermore, the like or dislike prediction accuracy was not correlated with sales of product

or self-reports in the study. By comparing with conventional marketing research methods such as sales or self-reports, the results of preference prediction based on EEG could be made

more reliable. More subjects would be recruited in the future to investigate the influence of some factors, such as brand and price, on the prediction results of consumer's preferences, while conducting correlation analysis with product sales and self-reports, etc.

CONCLUSION

This manuscript proposes a consumer's preference prediction system based on EEG by taking sport shoes as an example. The results show that the classification accuracy of 94.22% is achieved based on the DE features. The method proposed in this study can be used for product preference prediction. Furthermore, the number of EEG channels and recording location can be optimized to make the system easy-to-use and time-effective. In the future, the product like or dislike prediction results would be correlated with product sales and self-reports to make the results of EEG-based preference prediction more reliable.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

REFERENCES

- Aldayel, M., Ykhlef, M., and Al-Nafjan, A. (2020). Deep learning for eeg-based preference classification in neuromarketing. *Appl. Sci.* 10:1525. doi: 10.3390/app10041525
- Aldayel, M., Ykhlef, M., and Al-Nafjan, A. (2021). Recognition of consumer preference by analysis and classification EEG signals. *Front. Hum. Neurosci.* 14:604639. doi: 10.3389/fnhum.2020.604639
- Al-Nafjan, A., Hosny, M., Al-Wabil, A., and Al-Ohali, Y. (2017). Classification of human emotions from Electroencephalogram (EEG) signal using deep neural network. *Int. J. Adv. Comput. Sci. Appl.* 8, 419–425. doi: 10.14569/IJACSA.2017.080955
- Ariely, D., and Berns, G. S. (2010). Neuromarketing: the hope and hype of neuroimaging in business. *Nat. Rev. Neurosci.* 11, 284–292. doi: 10.1038/nrn2795
- Baldo, D., Parikh, H., Piu, Y., and Müller, K.-M. (2015). Brain waves predict success of new fashion products: a practical application for the footwear retailing industry. *J. Creat. Value* 1, 61–71. doi: 10.1177/2394964315569625
- Barnett, S. B., and Cerf, M. (2017). A ticket for your thoughts: method for predicting movie trailer recall and future ticket sales using neural similarity among moviegoers. *J. Consum. Res.* 44, 160–181. doi: 10.1093/jcr/ucw083
- Bazzani, A., Ravaioli, S., Trieste, L., Faraguna, U., and Turchetti, G. (2020). Is EEG suitable for marketing research? a systematic review. *Front. Neurosci.* 14:594566. doi: 10.3389/fnins.2020.594566
- Bell, L., Vogt, J., Willemse, C., Routledge, T., Butler, L. T., and Sakaki, M. (2018). Beyond self-report: a review of physiological and neuroscientific methods to investigate consumer behavior. *Front. Psychol.* 9:1655. doi: 10.3389/fpsyg.2018.01655
- Boksem, M. A. S., and Smidts, A. (2015). Brain responses to movie trailers predict individual preferences for movies and their population-wide commercial success. *J. Mark. Res.* 52, 482–492. doi: 10.1509/jmr.13.0572
- Cartocci, G., Caratù, M., Modica, E., Maglione, A. G., Rossi, D., Cherubino, P., et al. (2017). Electroencephalographic, heart rate, and galvanic skin response assessment for an advertising perception study: application to antismoking public service announcements. *J. Vis. Exp.* 126:55872. doi: 10.3791/55872

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Nanjing Brain Hospital Affiliated to Nanjing Medical University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LZ, HZ, and JW initiated and supervised the research project. LZ, ML, and KX carried out the research, analyzed the results and prepared the figures. LZ, ML, and HZ wrote part of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the National Natural Science Foundation of China (grant U1613228), Siyuan Foundation (grant HTKJ2020KL012002), and the open foundation of Key Laboratory of Biorheological Science and Technology (Chongqing University), Ministry of Education (grant CQKLBST-2020-005).

- Chen, D.-W., Miao, R., Yang, W.-Q., Liang, Y., Chen, H.-H., Huang, L., et al. (2019). A feature extraction method based on differential entropy and linear discriminant analysis for emotion recognition. *Sensors* 19:1631. doi: 10.3390/s19071631
- Golnar-Nik, P., Farashi, S., and Safari, M.-S. (2019). The application of EEG power for the prediction and interpretation of consumer decision-making: a neuromarketing study. *Physiol. Behav.* 207, 90–98. doi: 10.1016/j.physbeh.2019.04.025
- Goto, N., Lim, X. L., Shee, D., Hatano, A., Khong, K. W., Buratto, L. G., et al. (2019). Can brain waves really tell if a product will be purchased? inferring consumer preferences from single-item brain potentials. *Front. Integr. Neurosci.* 13:19. doi: 10.3389/fnint.2019.00019
- Hjorth, B. (1970). EEG analysis based on time domain properties. *Electroencephalogr. Clin. Neurophysiol.* 29, 306–310. doi: 10.1016/0013-4694(70)90143-4
- Joshi, V. M., and Ghongade, R. B. (2022). “EEG based emotion investigation from various brain region using deep learning algorithm,” in *ICDSMLA 2020, Lecture Notes in Electrical Engineering*, eds A. Kumar, S. Senatore, and V. K. Gunjan (Singapore: Springer), 395–402. doi: 10.1007/978-981-16-3690-5_34
- Khurana, V., Gahalawat, M., Kumar, P., Roy, P. P., Dogra, D. P., Scheme, E., et al. (2021). “A survey on neuromarketing using EEG signals,” in *Proceedings of the IEEE Transactions on Cognitive and Developmental Systems (Early Access)* (Piscataway, NJ: IEEE), 1–1. doi: 10.1109/TCDS.2021.3065200
- Khushaba, R. N., Greenacre, L., Kodagoda, S., Louviere, J., Burke, S., and Dissanayake, G. (2012). Choice modeling and the brain: a study on the Electroencephalogram (EEG) of preferences. *Expert Syst. Appl.* 39, 12378–12388. doi: 10.1016/j.eswa.2012.04.084
- Moon, J. (2013). Extraction of user preference for video stimuli using EEG-based user responses. *ETRI J.* 35, 1105–1114. doi: 10.4218/etrij.13.0113.0194
- Morin, C. (2011). Neuromarketing: the new science of consumer behavior. *Society* 48, 131–135. doi: 10.1007/s12115-010-9408-1
- Qing, K., Huang, R., and Hong, K.-S. (2021). Decoding three different preference levels of consumers using convolutional neural network: a functional near-infrared spectroscopy study. *Front. Hum. Neurosci.* 14:597864. doi: 10.3389/fnhum.2020.597864

- Ramsøy, T. Z., Skov, M., Christensen, M. K., and Stahlhut, C. (2018). Frontal brain asymmetry and willingness to pay. *Front. Neurosci.* 12:138. doi: 10.3389/fnins.2018.00138
- Sánchez-Fernández, J., Casado-Aranda, L.-A., and Bastidas-Manzano, A.-B. (2021). Consumer neuroscience techniques in advertising research: a bibliometric citation analysis. *Sustainability* 13:1589. doi: 10.3390/su13031589
- Shi, L.-C., Jiao, Y.-Y., and Lu, B.-L. (2013). "Differential entropy feature for EEG-based vigilance estimation," in *Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Osaka: IEEE), 6627–6630. doi: 10.1109/EMBC.2013.6611075
- Touchette, B., and Lee, S.-E. (2017). Measuring neural responses to apparel product attractiveness: an application of frontal asymmetry theory. *Cloth. Text. Res. J.* 35, 3–15. doi: 10.1177/0887302X16673157
- Venkatraman, V., Clithero, J. A., Fitzsimons, G. J., and Huettel, S. A. (2012). New scanner data for brand marketers: how neuroscience can help better understand differences in brand preferences. *J. Consum. Psychol.* 22, 143–153. doi: 10.1016/j.jcps.2011.11.008
- Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* 15, 70–73. doi: 10.1109/TAU.1967.1161901
- Yilmaz, B., Korkmaz, S., Arslan, D. B., Güngör, E., and Asyali, M. H. (2014). Like/dislike analysis using EEG: determination of most discriminative channels and frequencies. *Comput. Methods Programs Biomed.* 113, 705–713. doi: 10.1016/j.cmpb.2013.11.010

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.

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.

Copyright © 2022 Zeng, Lin, Xiao, Wang and Zhou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Target-Related Alpha Attenuation in a Brain-Computer Interface Rapid Serial Visual Presentation Calibration

Daniel Klee^{1*}, Tab Memmott^{1,2}, Niklas Smedemark-Margulies³, Basak Celik⁴, Deniz Erdogmus⁴ and Barry S. Oken^{1,5,6} on behalf of the Consortium for Accessible Multimodal Brain-Body Interfaces (CAMBI)

¹ Department of Neurology, Oregon Health and Science University, Portland, OR, United States, ² Institute on Development and Disability, Oregon Health and Science University, Portland, OR, United States, ³ Khoury College of Computer Science, Northeastern University, Boston, MA, United States, ⁴ Department of Electrical and Computer Engineering, Northeastern University, Boston, MA, United States, ⁵ Department of Behavioral Neuroscience, Oregon Health and Science University, Portland, OR, United States, ⁶ Department of Biomedical Engineering, Oregon Health and Science University, Portland, OR, United States

OPEN ACCESS

Edited by:

Ren Xu,
Guger Technologies, Austria

Reviewed by:

Long Chen,
Zhejiang Wolvo Bio-Pharmaceutical
Co. Ltd., China
Mengfan Li,
Hebei University of Technology, China
Manaf Altaleb,
University of Wasit, Iraq

*Correspondence:

Daniel Klee
klee@ohsu.edu

Specialty section:

This article was submitted to
Brain-Computer Interfaces,
a section of the journal
Frontiers in Human Neuroscience

Received: 23 February 2022

Accepted: 21 March 2022

Published: 21 April 2022

Citation:

Klee D, Memmott T, Smedemark-Margulies N, Celik B, Erdogmus D and Oken BS (2022) Target-Related Alpha Attenuation in a Brain-Computer Interface Rapid Serial Visual Presentation Calibration. *Front. Hum. Neurosci.* 16:882557. doi: 10.3389/fnhum.2022.882557

This study evaluated the feasibility of using occipitoparietal alpha activity to drive target/non-target classification in a brain-computer interface (BCI) for communication. EEG data were collected from 12 participants who completed BCI Rapid Serial Visual Presentation (RSVP) calibrations at two different presentation rates: 1 and 4 Hz. Attention-related changes in posterior alpha activity were compared to two event-related potentials (ERPs): N200 and P300. Machine learning approaches evaluated target/non-target classification accuracy using alpha activity. Results indicated significant alpha attenuation following target letters at both 1 and 4 Hz presentation rates, though this effect was significantly reduced in the 4 Hz condition. Target-related alpha attenuation was not correlated with coincident N200 or P300 target effects. Classification using posterior alpha activity was above chance and benefitted from individualized tuning procedures. These findings suggest that target-related posterior alpha attenuation is detectable in a BCI RSVP calibration and that this signal could be leveraged in machine learning algorithms used for RSVP or comparable attention-based BCI paradigms.

Keywords: electroencephalography (EEG), posterior alpha, attention, brain-computer interface (BCI), event-related potential (ERP), N200, P300, signal classification

INTRODUCTION

The development and application of brain-computer interface (BCI) technology has steadily increased over the past quarter-century (Wolpaw and Wolpaw, 2012; Rashid et al., 2020). Broadly, BCIs leverage neurophysiological signals to help users perform tasks related to movement or communication. BCI technology offers benefits to clinical populations for whom extant assistive technologies are insufficient, including individuals with locked-in syndrome (Wolpaw, 2013; Akcakaya et al., 2014). The scope, design, and neurophysiological mechanisms of BCI systems are quite varied, though BCI designs are popularly divided into two broad categories: (1) invasive systems that record data from intra-cranial electrodes, and (2) non-invasive systems that use scalp

electroencephalography (EEG). Non-invasive BCI systems are more popular because they are more affordable, accessible, and do not require surgery (Akçakaya et al., 2014; Rashid et al., 2020).

Recently, software tooling in Python has been made openly available to help facilitate BCI research and development. The open-source repository BciPy (Mémott et al., 2021) includes a common BCI paradigm for communication referred to as Rapid Serial Visual Presentation (RSVP; Huang et al., 2011; Orhan et al., 2012, 2013; Acqualagna and Blankertz, 2013; Lees et al., 2018). The RSVP task presents users with a sequence of images, such as letter characters, to assess target- and non-target-related brain responses derived from scalp EEG, including the event-related potentials (ERPs) N200 and P300, which indicate stimulus discrimination and attentional processing, respectively (Patel and Azzam, 2005). RSVP may be a particularly useful paradigm in cases where users are restricted by limited eye movement or visual attention (Fried-Oken et al., 2020), since RSVP only requires a user to fixate a single stimulus presented at the center of their gaze (Acqualagna and Blankertz, 2013; Akçakaya et al., 2014).

A brain-related oscillatory signature referred to as the alpha rhythm encompasses posterior-dominant activity in the range of approximately 8–13 Hz, though the exact bounds of alpha are often debated (Bazanov and Vernon, 2014). Alpha activity is understood to result from some thalamo-cortical synchronization occurring over many centimeters, rather than any localized cortical area (Lopes da Silva et al., 1980; Lopes da Silva, 2013). Of interest to the current work, changes in alpha activity have been experimentally linked to various cortical activations. Examples of these patterns include attenuation of central alpha or mu rhythm over regions secondary to actual or imagined contralateral limb movement (Pfurtscheller et al., 2006; Meirovitch et al., 2015), or attenuation of posterior alpha with eye opening or cognitive processing of visual stimuli (Klimesch, 1999; Pfurtscheller and Lopes da Silva, 1999). Additionally, changes in the distribution of posterior alpha are known to track shifts in covert spatial attention (Foster et al., 2017).

Prior research has explored the relationship between attention-related changes in alpha and ERPs. Some evidence has suggested that alpha desynchronization and target-related P300 may index and predict similar attentional processes (Yordanova et al., 2001; Grent-T-Jong et al., 2011), or that evoked potentials have the potential to interfere when measuring attention-related changes in alpha activity (van Gerven et al., 2009). However, it is known that visual attention can modulate alpha amplitude in the absence of ERP signals (Klimesch, 1999), and source analysis has associated these signals with different cortical structures (Peng et al., 2012).

A few studies have demonstrated the feasibility of using posterior alpha to track spatial attention in order to make letter selections in BCIs for communication (Kelly et al., 2005; van Gerven et al., 2009; van Gerven and Jensen, 2009; Treder et al., 2011). Evidence suggests that posterior alpha remains a viable signal for classification over repeated sessions (Horschig et al., 2015), and that patterns of alpha lateralization can be altered through neurofeedback training (Okazaki et al., 2015). In addition to spatial attention, posterior alpha might

also be used to track a user's mental state during BCI tasks (Myrden and Chau, 2017). However, despite being widely studied and relatively easy to measure, vision-modulated posterior alpha is neither a popular nor a common component of many BCI spellers (Rezeika et al., 2018; Rashid et al., 2020). To our knowledge, no previous work has attempted to use attention-related changes in posterior alpha to make letter selections in the context of a centrally fixated BCI spelling paradigm such as RSVP.

Given the evidence that visual attention modulates posterior alpha activity, and the conspicuous absence of posterior alpha as a driving neural signal in non-invasive BCI spelling systems, we conducted an exploratory study to examine the feasibility of using this signal in the context of the BciPy RSVP paradigm. A novel contribution of the current study is that no prior BCI-related work has examined posterior alpha changes outside of lateralized displays. Our primary aims were to determine whether target-related changes in posterior alpha activity are detectable in the RSVP task and, if so, whether these changes are sensitive to the presentation rate of letter stimuli. We hypothesized: (1) that event-related alpha activity would decrease following the presentation of target letter stimuli relative to non-target stimuli; and (2) that target-related posterior alpha attenuation would be smaller at an increased rate of presentation due to temporal overlap of target processing with subsequent non-targets. As secondary aims, we sought to assess whether target-related posterior alpha effects correlated with coincident target effects of N200 and P300 ERPs, and whether we could train machine learning algorithms to classify target and non-target responses using posterior alpha signals alone.

MATERIALS AND METHODS

EEG data were recorded from a convenience sample of generally healthy adults recruited at Oregon Health and Science University (OHSU) in Portland, OR. Research activities were registered with and approved by the OHSU Institutional Review Board (IRB). Data were collected over the course of a single 90-min session, after which participants were compensated \$25 for completing the study.

Participants

Demographic information is presented in **Table 1**. Twelve generally healthy adults enrolled in the study and provided written informed consent before participating in study activities, in accordance with the Declaration of Helsinki. All individuals recruited for the study were fluent in English. No participants reported use of alcohol or other mind-altering substances within 12 h of their test session. Exclusion criteria included use of EEG-altering medications such as neuroleptics or benzodiazepines (as reviewed by a physician), or an inability to perceive RSVP task stimuli and achieve at least 80% accuracy on a RSVP practice task. No one who enrolled was ineligible to participate based on these criteria. All participants were confirmed to have normal or corrected-to-normal vision and a minimum near-field Snellen visual acuity of 20/30 in at least one eye. All participants

TABLE 1 | Demographics.

	Participants (<i>n</i> = 12)
Age: mean years \pm SD (range)	33.75 \pm 6.40 (28–46)
Sex	
Female	6
Male	6
Race	
White	12
Ethnicity	
Hispanic/Latino	1
Not-Hispanic/Latino	11

A summary of participants enrolled in the study. Our sample contained a balanced number of female and male participants, but was homogeneous with regard to race and ethnicity.

achieved perfect scores on the practice RSVP task after only a single attempt.

Procedure

After providing written informed consent at the start of the study visit, participants self-reported demographic information followed by a brief health questionnaire and near-field visual acuity test. Upon completion of the screening questions, participants were introduced to a practice version of the RSVP calibration (see section “Rapid Serial Visual Presentation Practice”) before donning the EEG cap. Participants then completed two instances of RSVP calibration (see section “Rapid Serial Visual Presentation Calibration”). Individuals were asked to indicate how sleepy they felt immediately before, between, and after the two calibrations using the Stanford Sleepiness Scale (SSS; Herscovitch and Broughton, 1981; MacLean et al., 1992), since research has demonstrated that alpha activity fluctuates during periods of drowsiness (Cantero et al., 2002).

Screening

Self-Report Questionnaires

For screening purposes, individuals were asked to self-report current health conditions, medications, and provide details about their sleep habits. Specifically, participants were asked to indicate whether they had problems sleeping too much or too little, or if they had seen a physician for a sleep-related disorder.

Rapid Serial Visual Presentation Practice

To confirm participants’ ability to perceive the RSVP stimuli (see section “Rapid Serial Visual Presentation Calibration”), lab staff administered a practice version of the RSVP calibration created in PsychoPy3 Experiment Builder, version 3.0.0b11 (Peirce, 2007, 2009). Unlike the test version of RSVP calibration, the practice task presented only 10 trials of pseudo-random letter sequences, each of which had a 50% chance to contain the target letter. Stimuli in the letter sequences appeared at a rate of 4 Hz in order to emulate the more difficult experimental condition. Participants were asked after each 10-letter sequence to state whether the target had been present.

Rapid Serial Visual Presentation Calibration Overview

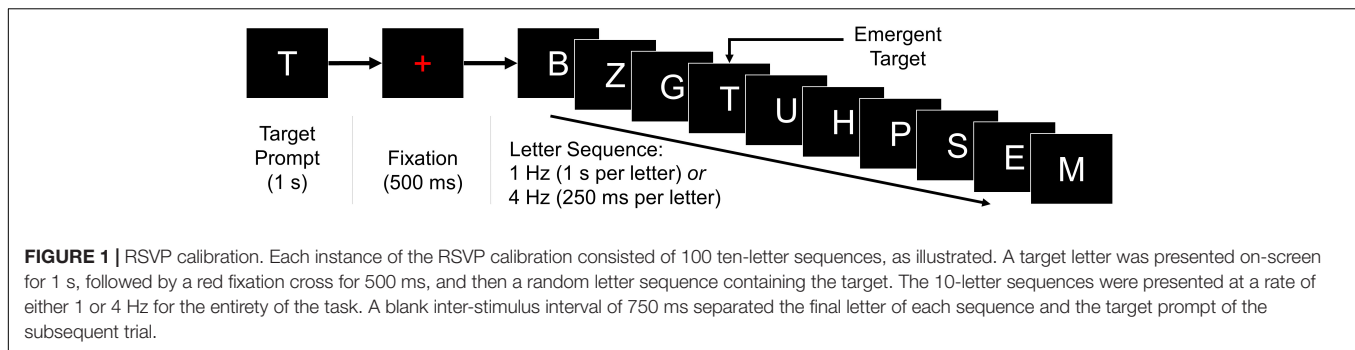
Participants completed two instances of the BciPy RSVP calibration (Orhan et al., 2011, 2012, 2013; Oken et al., 2014; Memmott et al., 2021). Each calibration consisted of 100 trials of 10-letter sequences and lasted approximately 10–25 min, depending on the presentation rate of letter stimuli. As demonstrated in **Figure 1**, each trial began with a target letter prompt presented for 1 s, followed immediately by the onset of a red fixation cross for 500 ms, and then a steady-stream sequence of 10 letter stimuli presented either at a rate of 1 or 4 Hz (1.0 s or 250 ms per stimulus, respectively). There were no blank-screen intervals within each sequence; a blank black screen was presented for 750 ms between sequences.

Participants were instructed to sit still, try to blink only between sequences, and watch for the target letter in each sequence, as indicated by the most recent target prompt. When the target appeared, participants were told to react mentally without movement or blinking. In cases where an individual was unable to resist blinking during the sequence (e.g., due to fatigue or irritation), there was a standing instruction to prioritize blinking during a non-target presentation in order to avoid missing the target. Along these lines, participants were allowed to pause the task by pushing the spacebar if they needed to rest.

Each participant completed RSVP calibrations with stimuli presented at rates of 1 Hz and 4 Hz; these rates were consistent within each 100-trial instance of the task. Condition order was balanced randomly across the 12 participant sessions in order to minimize systematic fatigue effects associated with repetition of the RSVP task (Oken et al., 2018). The logic behind our use of 1 and 4 Hz presentation rates was twofold: (1) to present stimuli at speeds where early visual and attentional processing of sequential stimuli would both overlap (i.e., 4 Hz; 250 ms per letter stimulus) and not overlap (i.e., 1 Hz; 1 s per letter stimulus); and (2) to minimize contamination of alpha-band activity by harmonics related to steady-state visual-evoked potential (SSVEP) artifact. Onset of the N200/P300 complex is expected 200–300 ms post-stimulus onset (Patel and Azzam, 2005), while target-induced posterior alpha desynchronization effects have been demonstrated approximately 300–800 ms following stimulus onset in a visual oddball paradigm (Peng et al., 2012). Past work with this RSVP paradigm has typically used a presentation rate of 5 Hz (e.g., Oken et al., 2018). However, as shown in **Figure 2**, the 2nd and 3rd harmonics of the SSVEP are sometimes apparent following spectral decomposition. To avoid contamination from a 10 Hz harmonic in the middle of the alpha band, the “typical” RSVP presentation rate of 5 Hz was lowered to 4 Hz.

Stimuli

RSVP stimuli were rendered on a 17.3-inch ASUS Vivobook Pro N705F laptop (1,920 \times 1,080 resolution) with a refresh rate of 60 Hz. Participants viewed the display from a seated position in a dimly lit room with consistent lighting. Viewing distance was approximately 70 cm, though head position was not constrained. Fixation crosses were drawn in solid red. Letter stimuli were



drawn in sans-serif Arial font and rendered solid white against an unchanging solid black background. Letter stimuli included 28 possible characters: capitalized forms of all 26 letters of the English alphabet, plus the characters “_” and “<,” which represent “space” and “backspace” in this RSVP paradigm, respectively.

At a viewing distance of 70 cm, the red fixation cross subtended 2.5° visual angle with a stroke width of 0.4°. All letter stimuli had a stroke width of approximately 0.5°. Due to morphological differences in the 28 letter stimuli, character height ranged between 0.5° (i.e., “_”) and 3.8° (e.g., “T”). Similarly, letter width ranged between 0.5° (i.e., “I”) and 4.8° (i.e., “W”).

Electrophysiological Recordings

EEG data were recorded from a DSI VR-300 headset (Wearable Sensing; San Diego, CA, United States). This adjustable dry-electrode system utilized a linked-ear reference with a ground at A1. The sampling frequency was 300 Hz, with an A/D resolution of 16 bits. Electrodes were placed according to a custom montage of standard 10/20 sites: FCz, F7, Pz, P4, PO7, PO8, and Oz. These sites are optimized to capture P300 evoked potentials (Krusienski et al., 2008), with the alteration that site P3 was replaced by F7 as an additional anterior site in order to capture eye-related activity (e.g., blinking). Signal quality was assessed in Wearable Sensing’s DSI-Streamer software (v.1.08.44) prior to completion of the experimental tasks. Electrodes were adjusted to satisfy manufacturer-determined default operating thresholds pertaining to clipping, noise (very low < 0.07 μ Vrms; very high > 100 μ Vrms), impedance (<2.0 M Ω), impedance signal-to-noise (<3.0 M Ω), and DC offset (50% of maximum). All experimental EEG data were recorded using the BciPy acquisition client.

Electrophysiological Processing

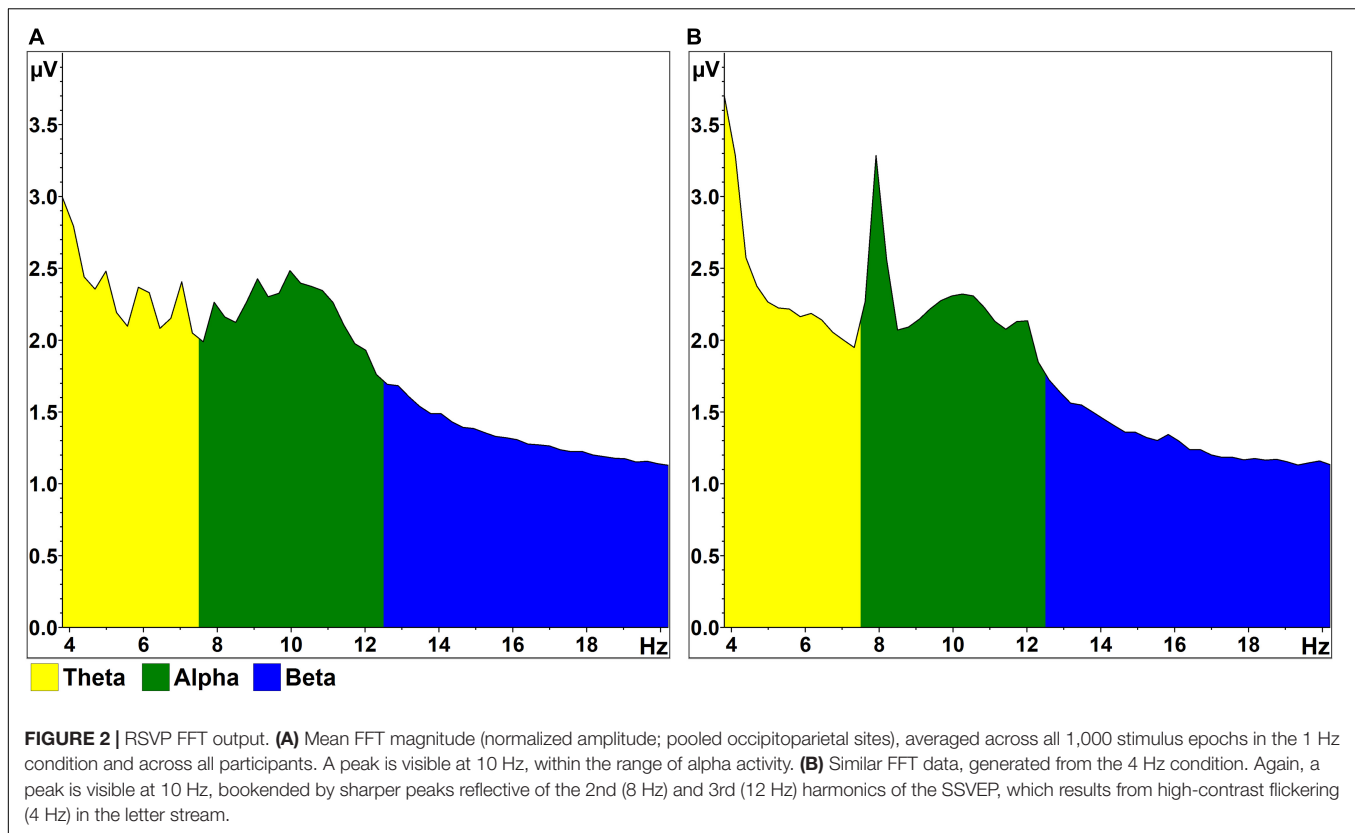
With the exception of the secondary BCI classification comparisons (see section “Brain-Computer Interface Classifiers”), all time-frequency and ERP analyses were performed offline in BrainVision Analyzer: Professional Edition, version 2.1.0.327 (BrainVision LLC; Morrisville, NC, United States). Due to *a priori* uncertainty regarding the exact distribution of the attention effect across the scalp during RSVP calibration, EEG measures were analyzed twice in parallel: both as a “pooled” (i.e., averaged) signal across occipitoparietal sites Pz, Oz, PO7, and PO8, and also at site Pz in isolation

(i.e., “Pz-only”). Pz is typically where experimenters can expect to observe the highest-amplitude target-related P300 (Polich, 2007), and was therefore selected as a representative sub-site. Data were filtered 1–45 Hz (48 dB/octave) along with a 60 Hz notch filter using a Butterworth zero-phase infinite impulse response (IIR) filter. These filtered data were then downsampled from 300 to 150 Hz.

Time-Frequency Analyses

EEG data in the time-frequency analysis were segmented into 2.5 s epochs centered relative to stimulus onset, such that each of the 100 target and 900 non-target epochs in a given calibration ranged from −1,250 ms to +1,250 ms relative to stimulus onset. This window was chosen in order to capture activity ± 1 s relative to stimulus onset, with ample buffer length of 250 ms at either end of the window to avoid edge-related artifact due to the wavelet analysis. Within each calibration recording, epochs were separated by condition and converted to time-frequency scaleograms using a continuous wavelet transform (CWT) with a complex-valued Morlet mother wavelet (Morlet parameter $c = 5$). Scaleograms were generated from wavelets with scaled frequencies ranging from 4 to 16 Hz in 48 logarithmic steps, normalized according to uniform scale power (unit energy normalization; all scaled frequency layers of the wavelet function possessed an energy value of 1). Complex voltage output was converted to real-valued voltage (μ V). The mean and standard deviation of a 500 ms baseline window ranging from −600 to −100 ms before stimulus onset were used to Z-transform all output samples within the test epochs. Unless otherwise specified, “alpha activity” in this analysis refers to the average of the Z-scored real voltage values within the designated response window of 300–800 ms post-stimulus onset, which was selected to capture the onset and duration of the target effect as seen in previous research with visual oddball paradigms (Peng et al., 2012; Vázquez-Marrufo et al., 2019). In other words, “negative” activity values indicate event-related decreases in alpha activity relative to the baseline window (−600 to −100 ms). See BrainVision Analyzer User Manual (Software Version 2.1.0) for further information regarding operations (Brain Products, 2014).

Our analyses examined alpha activity following target and non-target letter stimuli, as well as the alpha attenuation effect, which we defined as the difference between target and non-target alpha responses within each RSVP calibration. Statistical tests were performed on alpha activity estimates taken from



the single wavelet layer with scaled central frequency closest to each participant's individual alpha frequency (IAF) within a calibration, rounded to the nearest 0.5 Hz. Identifying spectral prominences prior to time-frequency investigations is a recommended technique to avoid misleading analyses of arbitrary noise (Pfurtscheller and Lopes da Silva, 1999; Corcoran et al., 2018). IAF estimates were determined from mean fast Fourier transformation (FFT) output from the pooled occipitoparietal signals; this FFT output was averaged across all available 2.5 s epochs within each individual instance of RSVP calibration (**Figure 2**). Real-valued voltage output was generated using a 20% Hanning window, periodic variance correction, and amplitude normalization relative to a frequency bandwidth of 4–20 Hz. Semi-automatic peak detection identified the frequency in the range of 7.5–12.5 Hz with the highest magnitude. Estimates were then reviewed visually and adjusted in the event of peak capture by either noise or SSVEP artifact in the 4 Hz presentation condition. Clear alpha peaks were absent in two of the 1 Hz calibrations and three of the 4 Hz calibrations. In these cases where no discernable peaks were present, the peak estimate was set manually to 10 Hz. The underlying rationale of this approach was that a wavelet layer with scaled central frequency of 10 Hz and spectral bandwidth of approximately 4 Hz is well-positioned to capture most of the signal within the alpha band (8–12 Hz). Peak values determined in this manner ranged from 9 to 11.5 Hz across participants; all participants demonstrated similar IAF estimates between the 1 and 4 Hz presentation conditions, with the exception of three individuals who showed marginal increases

in peak alpha of 0.5 Hz (two instances) and 1.0 Hz (one instance) in the 4 Hz condition.

Event-Related Potential Analyses

To accommodate ERP analyses, the original data were segmented into 1 s epochs ranging from –200 to +800 ms, relative to stimulus onset. These epochs were baseline corrected (–200 to 0 ms), separated by stimulus-type, and averaged within-condition before the use of semi-automatic peak detection to label N200 (200–350 ms) and P300 (300–450 ms) potentials in the target condition. N200 and P300 responses within both the target and non-target classes were estimated as mean voltage (μV) \pm 4 sampled points (~ 53 ms) around the labeled peak latencies derived from the target samples. Our analyses measured N200 and P300 signed voltages following target and non-target letter stimuli, as well as the target effect, which we defined as the difference between target and non-target voltage responses within each condition.

Artifact Rejection

To approximate the processing tools included by default in BciPy, which currently does not include tooling for artifact handling outside of filtering, our primary offline analyses of EEG and ERP data did not implement rigorous artifact rejection procedures. However, to ensure that our results were not influenced by poor data quality and that electrooculographic (EOG) activity had no effect on our alpha measurements, we re-ran key analyses using data that included artifact rejection to remove eye blinks,

extraneous electromyography (EMG), and other non-descript noise such as electrode popping due to movement and poor electrode contact. All time-frequency and ERP epochs were flagged for review and semi-automatic rejection according to the following rules: voltage gradients $> 50 \mu\text{V}/\text{ms}$; voltage differences $> 125 \mu\text{V}$ within a window of 50 ms; absolute voltage values $> 75 \mu\text{V}$; and activity $< 0.5 \mu\text{V}$ sustained for ≥ 100 ms in duration. Data marked in this manner were reviewed visually and contaminated epochs were discarded prior to separating epochs by stimulus class (i.e., target vs. non-target).

Artifact rejection results are discussed under “Results” section “Artifact Rejection.” Individual participants kept at least 70% of target time-frequency epochs in the 1 Hz condition and 65% of targets at 4 Hz, with the exceptions that one participant only maintained 59 target segments in the 1 Hz condition, and another kept only 46 target segments in the 4 Hz condition. Across ERP epochs, individual participants kept $\geq 85\%$ of target segments at 1 Hz and $\geq 78\%$ of targets at 4 Hz.

Brain-Computer Interface Classifiers

The Python code underlying our BCI classifier analyses is accessible online (Zenodo, 2022).

Brain-Computer Interface Classifiers: Alpha Data

For the purpose of training classifiers using alpha-band features, we performed analogous time-frequency preprocessing of the raw EEG data using the PyWavelets Python package (Lee et al., 2019). Due to differences in software, the wavelet parameters used here differed slightly from those outlined above in section “Time-Frequency Analyses”: we used a Morlet mother wavelet with bandwidth of 1.5 Hz and a central frequency of 1.0 Hz; scaled wavelets were not normalized. Data were Z-scored according to the same 500 ms baseline (−600 to −100 ms) and response (+300 to +800 ms) windows, and we selected one frequency to investigate for each participant (see IAF in “Materials and Methods” section “Time-Frequency Analyses”). Unlike the primary analysis, however, alpha activity was preserved as time-series data and was not averaged within the response window. We also experimented with optimizing the locations of these 500 ms baseline and response windows to improve classification. We varied the baseline window start between −1,050 and −600 ms, and varied the response window start between +150 and +550 ms, selecting the best starting positions according to the balanced test accuracy of a Logistic Regression classifier. Classifier analyses were conducted using data from all four of the occipitoparietal channels.

We evaluated several classifiers on this time-series representation of the alpha data, including Logistic Regression with L2 regularization and a support vector classifier. We also tried classifying these data in a channelwise covariance matrix representation, using a logistic regression classifier on a tangent space projection (Barachant et al., 2012, 2013). The experimental setup results in a fixed class ratio of 9 non-targets for every target; thus, all models were trained using class-balanced objectives and evaluated using balanced accuracy, which is the average of positive (target) and negative (non-target) class accuracy: $acc_{bal} = (acc_{target} + acc_{non-target})/2$. Note that a model that

makes uniform random guesses will achieve a balanced accuracy of approximately 0.5 (i.e., at chance levels), while a perfect model will achieve an accuracy of 1.0.

Brain-Computer Interface Classifiers: Event-Related Potential Data

To compare alpha classification performance to the default classification mechanism used in BciPy, we used principal component analysis, followed by regularized discriminant analysis and kernel density estimation (RDA/KDE) to classify filtered EEG time-series data inclusive of the N200 and P300 ERPs. This approach has been detailed in previous work (Oken et al., 2018; Memmott et al., 2021) and was used to generate balanced accuracy estimates similar to those of the alpha classifiers outlined in section “Brain-Computer Interface Classifiers: Alpha Data.” Unlike the ERP analyses in “Materials and Methods” section “Event-Related Potential Analyses,” ERP time-series data were filtered 2–45 Hz and segmented into 500 ms epochs ranging 0 to +500 ms relative to stimulus onset in order to match the default settings of BciPy.

Statistical Analyses

Statistical tests were conducted in IBM SPSS Statistics, Version 27 (IBM Corporation; Armonk, NY, United States). Shapiro-Wilk tests were used to assess the normality of our measures; box plots were reviewed to identify relevant outliers. Roughly one-third of the primary across-participant alpha and ERP measures were not normally distributed, trending ($p < 0.10$) toward non-normality, or contained at least one outlier. Because these measures were inconsistently normally distributed, we opted to forego parametric means comparisons and instead conducted across-participant median comparisons using non-parametric two-tailed related-samples Wilcoxon signed rank tests. Shapiro-Wilk tests also revealed that most of the within-participant alpha measures were not normally distributed, so independent-samples Mann-Whitney *U*-tests were used to evaluate target vs. non-target alpha activity differences within individual RSVP calibrations. These data were inspected visually to confirm general similarity of shape. Spearman’s rank-order correlations were used to examine the relationships between alpha attenuation effects and ERP target effects. Similarly, artifact rejection and tuned parameter analyses used non-parametric tests to accommodate non-normal data and outliers.

RESULTS

Self-reported sleepiness levels demonstrated a significant median increase between the start ($Mdn = 3.0$) and mid-point ($Mdn = 3.5$) of the experiment, $Z = -2.111$, $p = 0.035$. However, there were no significant median differences in sleepiness levels before RSVP, after RSVP, or sleepiness changes after-minus-before RSVP according to condition, either 1 or 4 Hz presentations (all p values ≥ 0.165). None of the self-reported SSS sleepiness metrics correlated with any of the alpha attenuation effects [all absolute $r_s(10)$ values ≤ 0.511 ; all p values ≥ 0.09].

Time-Frequency Analyses

Time-frequency scaleograms displayed in **Figure 3** show representative wavelet output 4–16 Hz during the RSVP calibrations. As clarified previously in section “Artifact Rejection,” these primary analyses included all 1,000 available epochs (both target and non-target) from all 100 trials in each RSVP calibration. Artifact rejection comparisons and the corresponding redundant analyses are discussed separately below in “Results” section “Artifact Rejection.”

Alpha Effects: Across-Participants

Figure 4 offers a summary of Z-scored alpha activity waveforms across participants (\pm standard error of the mean; SEM) according to electrode source, presentation rate, and stimulus class. Measuring the pooled occipitoparietal signal in the 1 Hz condition, target-related median alpha activity ($Mdn = -0.164$) was significantly lower than for non-target letters ($Mdn = 0.285$), $Z = -2.667$, $p = 0.008$. Similarly, in the 4 Hz condition, median alpha activity was lower following targets ($Mdn = -0.022$) compared to non-targets ($Mdn = 0.172$), but this difference did not reach statistical significance, $Z = -1.883$, $p = 0.060$. The size of the target-vs.-non-target attenuation effect was significantly larger at 1 Hz presentation rates ($Mdn = -0.632$) compared to 4 Hz presentation rates ($Mdn = -0.227$), $Z = -2.353$, $p = 0.019$.

Using data from Pz-only, we observed effects similar to those in the pooled signal. In the 1 Hz condition, median alpha activity for targets ($Mdn = -0.103$) was significantly lower than median alpha related to non-targets ($Mdn = 0.373$), $Z = -3.059$, $p = 0.002$. At 4 Hz presentations, median target-related alpha activity ($Mdn = 0.098$) was again lower than that of the non-target stimuli ($Mdn = 0.196$), though unlike data from the pooled signal, this difference was significant at Pz-only, $Z = -1.961$, $p = 0.050$. The median alpha attenuation effect at Pz-only was larger in the 1 Hz condition ($Mdn = -0.527$) compared to the 4 Hz condition ($Mdn = -0.103$), $Z = -2.510$, $p = 0.012$. There were no meaningful differences between median alpha attenuation in the pooled signal and at Pz-only, either at 1 or at 4 Hz presentations (both comparisons: $Z = -0.784$, $p = 0.433$).

Alpha Effects: Within-Participants

While between-participant effects are of interest, the use of alpha signals within BCI systems ultimately requires significant within-participant effects. Independent-samples Mann-Whitney U comparisons revealed significant median differences between target and non-target alpha Z-scores within individual RSVP calibrations. Significant attenuation of alpha activity following target letter stimuli was discernable in at least one of the signal sources (i.e., either the pooled signal or Pz-only) in 9/12 participants in the 1 Hz condition, and 6/12 participants in the 4 Hz condition (see **Figure 5** for a summary of within-participant alpha attenuation effects). The attenuation effect was equally detectable between the pooled occipitoparietal signal and Pz-only during 1 Hz presentations, such that 7 participants showed significant median differences between target and non-target alpha activity in both signal sources, while 1 participant evinced the attenuation effect only in the pooled signal and another just at Pz-only. This pattern was comparable in the 4 Hz condition,

where 3 participants demonstrated the effect in both signals, 2 showed changes only in the pooled signal, and 1 just at Pz-only.

Event-Related Potential Analyses

Many of the expected ERP target effects were observed across pooled occipitoparietal sites (**Figure 6**). N200 and P300 measures were all increased following target stimuli compared to non-target stimuli, both at 1 and 4 Hz presentation rates (all p values ≤ 0.012). The size of the N200 target effect was smaller in the 1 Hz condition ($Mdn = -1.629$) compared to the 4 Hz condition ($Mdn = -3.842$), $Z = -2.746$, $p = 0.006$, while the corresponding P300 target effect was inversely greater in the 1 Hz condition ($Mdn = 3.262$) than in the 4 Hz condition ($Mdn = 1.508$), $Z = -2.435$, $p = 0.015$. These ERP target effects remained unchanged when measured at Pz-only, with increased N200 and P300 measures for targets compared to non-targets (all p values ≤ 0.003). The median N200 target effect was larger during 4 Hz presentations ($Mdn = -3.512$) compared to 1 Hz presentations ($Mdn = -2.648$), $Z = -2.746$, $p = 0.006$, and the median P300 target effect was again larger at 1 Hz ($Mdn = 5.513$) compared to the 4 Hz condition ($Mdn = 3.501$), $Z = -2.118$, $p = 0.034$. N200 target effects were not discernably different between the pooled signals and Pz-only in either presentation condition (both p -values ≥ 0.433). P300 target effects were larger at Pz-only than in the pooled signal in both the 1 and 4 Hz conditions (both p values = 0.002).

Correlation of Across-Participant Alpha Attenuation and Event-Related Potential Target Effects

Alpha attenuation and ERP target effects are depicted side-by-side as difference waveforms in **Figure 7**. Across pooled occipitoparietal sites in the 1 Hz condition, alpha attenuation effects did not correlate with N200 [$r_s(10) = -0.133$, $p = 0.681$] or P300 target effects [$r_s(10) = -0.105$, $p = 0.746$]. Pooled signal alpha attenuation effects in the 4 Hz condition did not meaningfully predict N200 [$r_s(10) = -0.329$, $p = 0.297$] or P300 target effects [$r_s(10) = -0.007$, $p = 0.983$]. This disconnect between alpha and ERP target effects in the pooled occipitoparietal signal was also evident in the Pz-only signal. As before, in the 1 Hz condition there were no significant correlations between alpha attenuation and ERP target effects for N200 [$r_s(10) = 0.084$, $p = 0.795$] or P300 [$r_s(10) = 0.035$, $p = 0.914$]. At 4 Hz, the Pz-only signal did not yield any significant correlations between alpha attenuation and target effects for either N200 [$r_s(10) = -0.042$, $p = 0.897$] or P300 [$r_s(10) = 0.000$, $p = 1.000$].

Artifact Rejection

Related-sample Wilcoxon signed ranks tests indicated that none of the between-participant alpha measures were statistically different before and after artifact rejection, as outlined in “Materials and Methods” section “Artifact Rejection” (all p values > 0.05). The results of our alpha comparisons remained unchanged, with a single exception that the difference between median target and non-target alpha activity at Pz-only was no

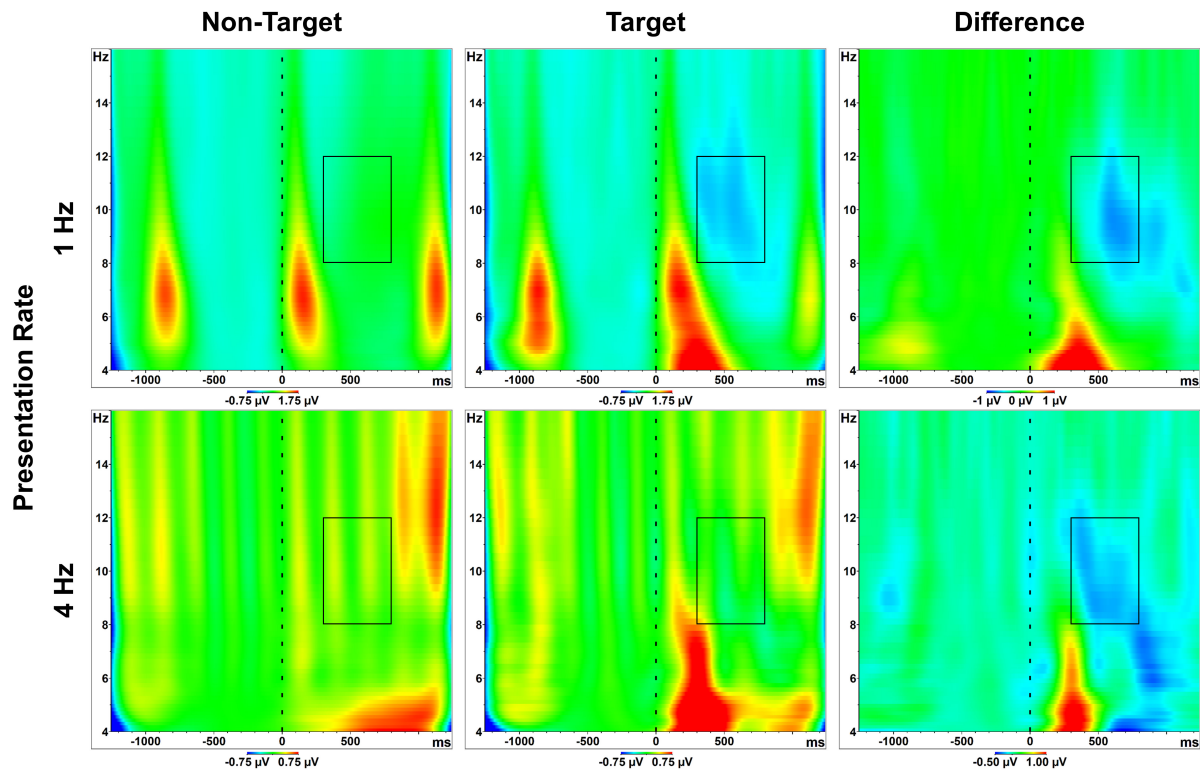


FIGURE 3 | Scaleograms. Spectra 4–16 Hz, relative to stimulus onset ($t = 0$). Plots are derived from pooled posterior sites, averaged across all participants, and separated by condition (presentation rate; stimulus class). Boxes represent the alpha band (8–12 Hz) and the measured effect window. The final column illustrates the alpha attenuation effect, or the difference in activity between target and non-target letter responses. SSVEP signatures are visible across conditions, both as conspicuous bursts in the range of 5–9 Hz (the 1 Hz presentation condition), and as more spectrally-diffuse vertical stripes (the 4 Hz presentation condition). A clear increase in theta activity is also visible approximately 200–300 ms following target stimuli. Color scaling is consistent between targets/non-targets within presentation rate conditions. However, due to gross differences in alpha activity between the 1 Hz and 4 Hz conditions (see **Figure 4**, which includes static scaling of axes across conditions), scaling was adjusted to better highlight the time-course of the attenuation effect.

longer statistically different at 4 Hz presentations ($p = 0.308$). Spearman's rank-order correlations indicated no significant correlations between alpha attenuation and ERP target effects in either the pooled [all absolute $r_s(10)$ values ≤ 0.476 ; all p values ≥ 0.118] or Pz-only signals [all absolute $r_s(10)$ values ≤ 0.238 ; all p values ≥ 0.457].

Within-participants, artifact rejection did not result in visible changes to the shape of alpha activity distributions; the majority of measures remained non-normal, including at least one class (i.e., target or non-target) within each electrode source per individual RSVP calibration. Artifact rejection resulted in changes to the alpha attenuation effect primarily in the 4 Hz condition, where 3 previously significant effects dissipated and 1 moved from significant ($p < 0.05$) to trending ($p < 0.10$). In the 1 Hz condition as well, 1 previously significant effect shifted to trending while another dissipated entirely.

Implementation and Classification

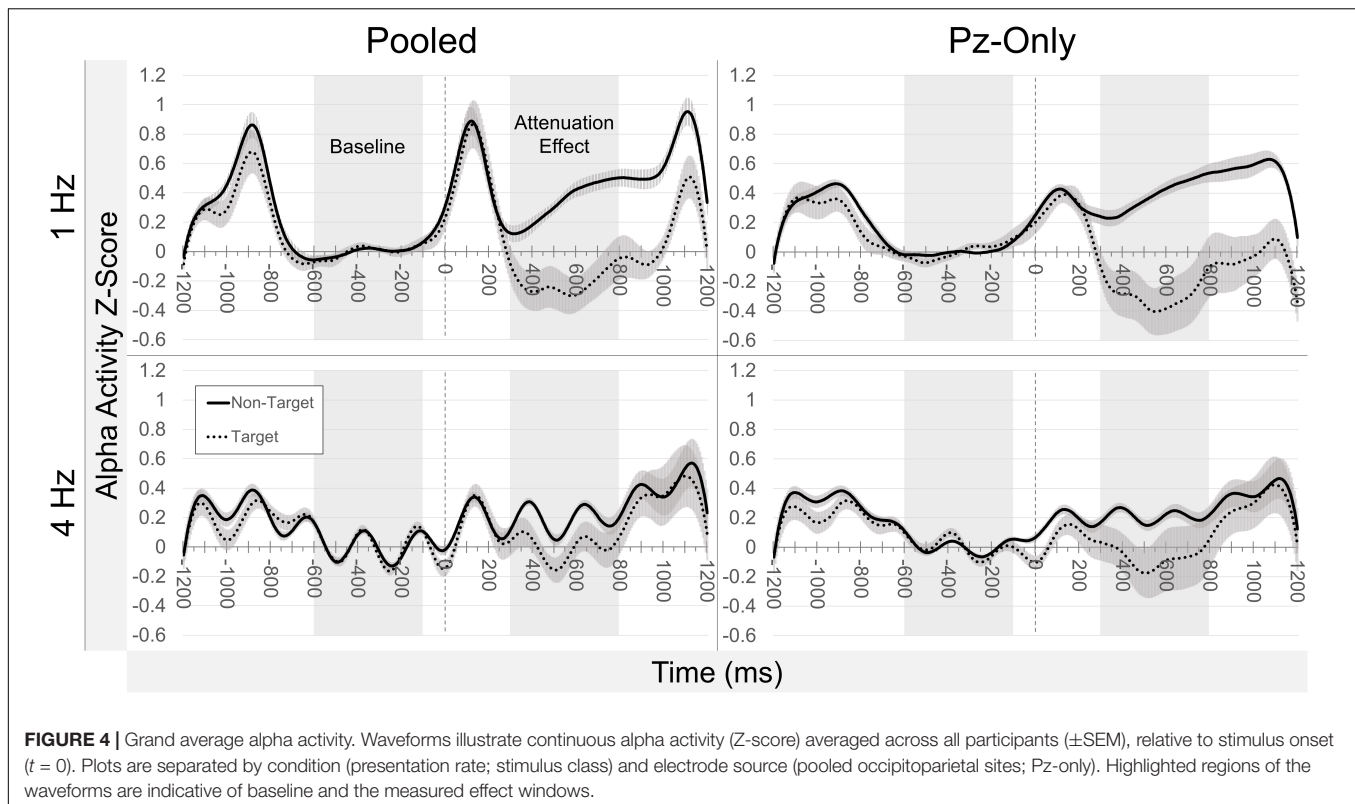
Regularized Discriminant Analysis and Kernel Density Estimation

Mean balanced test accuracies from the ERP RDA/KDE model ranged 0.595–0.836 across participants in the 1 Hz condition and

0.614–0.874 in the 4 Hz condition. A paired-samples Wilcoxon test indicated a significant median increase in test accuracies between the 1 Hz condition ($Mdn = 0.676$) and the 4 Hz condition ($Mdn = 0.707$), $Z = -2.433$, $p = 0.015$. Within the 1 Hz condition, there was a significant Spearman's correlation between mean balanced test accuracies of the RDA/KDE model and N200 target effects [$r_s(10) = -0.776$, $p = 0.003$]. The same RDA/KDE estimates related to P300 target effects [$r_s(10) = 0.531$, $p = 0.075$], but only weakly. 1 Hz condition RDA/KDE accuracies were not correlated with alpha attenuation [$r_s(10) = 0.287$, $p = 0.366$]. At 4 Hz presentations, mean balanced test accuracies from the RDA/KDE model were associated with N200 [$r_s(10) = -0.790$, $p = 0.002$] and P300 target effects [$r_s(10) = 0.608$, $p = 0.036$], though RDA/KDE accuracies were again unrelated to alpha attenuation effects [$r_s(10) = 0.077$, $p = 0.812$].

Alpha Classifiers

A comparative bar chart of the various alpha and ERP classifier methods is shown in **Figure 8**. Default parameters yielded mean balanced test accuracies for individual participants that ranged 0.470–0.764 across models in the 1 Hz condition and 0.463–0.729 in the 4 Hz condition. Tuned parameters resulted



in mean balanced test accuracies ranging 0.508–0.772 across models in the 1 Hz condition and 0.472–0.732 in the 4 Hz condition. Across participants at both 1 and 4 Hz presentations, both with default and tuned parameters, all alpha modeling underperformed relative to accuracy estimates from the ERP RDA/KDE model (Wilcoxon: all p values ≤ 0.008). Alpha attenuation effects were significantly correlated with all default classifier mean test accuracy estimates at 1 Hz (p values ≤ 0.039), with the expected exception that there was no meaningful correlation between alpha attenuation and the Uniform Random model. In contrast, alpha attenuation in the 4 Hz condition was largely unrelated to accuracy estimates from the default classifiers (p values ≥ 0.249), except for a weak correlation between alpha and Logistic Regression model estimates [$r_s(10) = -0.515$, $p = 0.087$].

All alpha models using default window parameters in the 1 Hz condition yielded mean test accuracies with a statistically higher median than chance (Uniform Random model, $Mdn = 0.509$; all p values ≤ 0.005). However, these relationships were much weaker during 4 Hz presentations, such that none of the three alpha classifiers were statistically distinguishable from the Uniform Random model ($Mdn = 0.481$; p values ≥ 0.06). Using individually tuned baseline and effect window latencies, all alpha models performed above chance ($Mdn = 0.514$) in the 1 Hz condition (all p values ≤ 0.005). In the 4 Hz condition however, Tangent Space ($Mdn = 0.523$) accuracies fell to the level of the Uniform Random model ($Mdn = 0.501$; $p = 0.126$), though the Logistic Regression ($Mdn = 0.542$) and Support Vector ($Mdn = 0.533$) models performed above

chance (p values ≤ 0.050). Expectedly, tuned classifier mean test accuracies demonstrated higher median values for Logistic Regression models in both the 1 Hz ($p = 0.045$) and 4 Hz conditions ($p = 0.005$), relative to their default-parameter counterparts. Parameter tuning had no significant effect on the accuracy estimates of the Uniform Random, Support Vector, or Tangent Space models.

Supplementary Analyses

Using tuned time-frequency analysis parameters, we reprocessed and reanalyzed the alpha data in BrainVision Analyzer in order to describe changes due to altered baseline and effect windows. Because tuning was performed using all four occipitoparietal channels, the supplementary across-participant analysis only utilized the pooled signal. Alpha activity distributions were normally distributed across participants. However, to remain consistent with our other analyses, median comparisons were again run using related-samples Wilcoxon signed-ranks tests. Median target ($Mdn = -0.179$) alpha activity was lower than median non-target activity ($Mdn = 0.284$) at 1 Hz presentations, and the same was true for targets ($Mdn = 0.001$) and non-targets ($Mdn = 0.187$) in the 4 Hz condition (both comparisons: $Z = -2.589$, $p = 0.010$). Alpha attenuation effects measured after parameter tuning did not correlate with any of the ERP target effects [all absolute $r_s(10)$ values ≤ 0.336 ; all p values ≥ 0.286].

Much like the within-participant alpha distributions discussed in sections “Alpha Effects: Within-Participants” and “Artifact Rejection,” distributions generated from individually tuned parameters remained non-normal in shape. Three previously

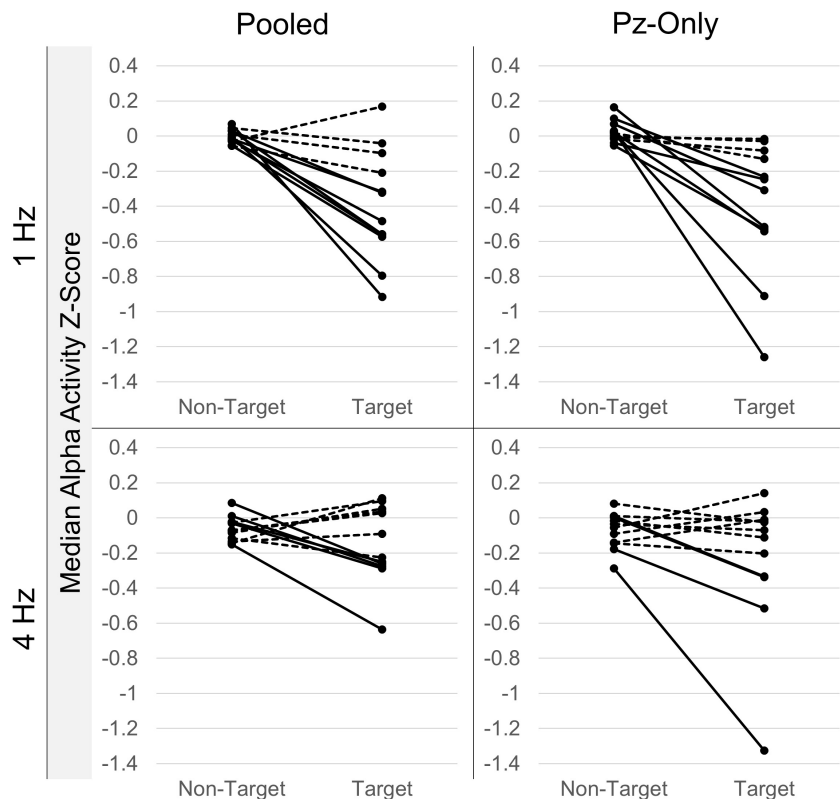


FIGURE 5 | Intra-individual alpha attenuation effects. Median target and non-target alpha activity Z-scores from individual RSVP calibrations. Data are sorted according to presentation rate and electrode source. Lines connect corresponding non-target and target estimates from within-individual recordings. Significant within-participant non-target/target median differences ($p < 0.05$; Mann-Whitney U -tests) are denoted by solid lines; non-significant differences ($p > 0.05$) are drawn as dashed lines. Plots show clear variability in the attenuation effect across individual participants and presentation conditions. Alpha attenuation effects appear to be greater at Pz for select individuals, but also more variable across participants.

significant alpha attenuation effects weakened to a trending state ($p < 0.10$), while another 3 were no longer significant following tuning procedures. However, two previously not-significant effects shifted to a trending state (note: one of these was in the opposite direction), while another became significant. Interestingly, the one effect that moved from not-significant to significant following tuning was observed in an individual with one of the weaker alpha attention effects, as measured in the primary analysis. Overall, following tuning procedures in the 1 Hz condition, 7/12 individuals demonstrated alpha attenuation in at least one signal source (6 at both pooled and Pz-only signals; 1 only in the pooled signal), with an additional 3 trending (2 in the pooled signal; 1 at Pz-only). One-half (6/12) of participants demonstrated alpha attenuation in at least one tuned signal source during the 4 Hz condition (1 in both sources; 3 only in the pooled signal; 2 at Pz-only).

DISCUSSION

The primary objectives of this exploratory study were (1) to detect target-related posterior alpha attenuation in a BCI RSVP paradigm, and (2) to discern whether that attenuation effect varied according to the presentation rate of the stimuli in

that task. Secondly, this study attempted to describe the relationship between changes in alpha and coincident ERP signals, and explore a subset of classification approaches to distinguish target and non-target alpha responses.

Summary of Findings

As hypothesized, we observed significant event-related attenuation of posterior alpha activity across participants for target letter stimuli relative to non-targets. This alpha attenuation effect was observed in both 1 and 4 Hz presentation conditions, though as expected, target-related attenuation was significantly greater in the 1 Hz condition. Within-participant findings were similar, such that more individuals demonstrated target-related posterior alpha attenuation in the 1 Hz condition than in the 4 Hz condition. Significant target effects were observed for both N200 and P300 ERPs, but were unrelated to alpha attenuation. Classifier models trained to target/non-target classification of alpha activity performed above chance, but underperformed markedly relative to RDA/KDE estimates generated from ERP time-series data. Alpha attenuation effects were observed using Z-transforms based on pre-defined baseline and response windows. Individualized parameter tuning of these temporal windows resulted in improved performance of one of the alpha classifiers.

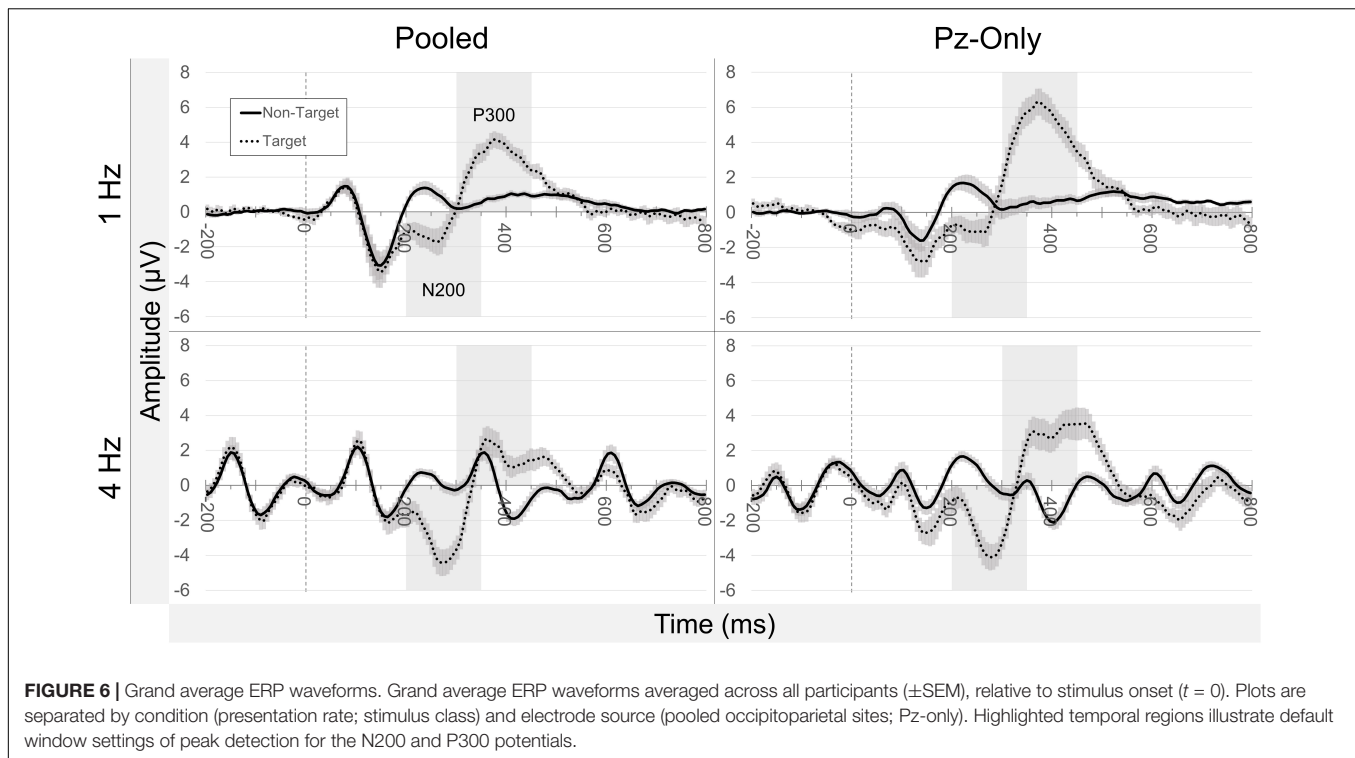


FIGURE 6 | Grand average ERP waveforms. Grand average ERP waveforms averaged across all participants (\pm SEM), relative to stimulus onset ($t = 0$). Plots are separated by condition (presentation rate; stimulus class) and electrode source (pooled occipitoparietal sites; Pz-only). Highlighted temporal regions illustrate default window settings of peak detection for the N200 and P300 potentials.

The Disconnect Between Alpha Activity and Event-Related Potentials

In review, there is evidence that (1) posterior alpha and ERPs such as the P300 can capture similar attentional processes (Yordanova et al., 2001; Grent-T-Jong et al., 2011), but also, that (2) alpha activity and attentional ERPs are not necessarily identical neural indices of attention (Klimesch, 1999; Peng et al., 2012). In agreement with these previous findings, the non-overlapping nature of alpha, N200, and P300 activity was apparent in the current study. Specifically, target-related fluctuations in alpha activity did not correspond to similar target-related changes in N200 or P300 signals across individuals.

Individual Differences

Although posterior alpha attenuation was visible across-participants, within-participant analyses revealed clear individual differences in both the presence and size of the alpha attenuation effect. Individual differences in alpha activity are well-documented, including differences in IAF (Corcoran et al., 2018), as well as the amount and distribution of alpha activity (Bazanov and Vernon, 2014). With these differences in mind, it seems fair to question the absolute utility of posterior alpha for BCI target classification across all individuals. As shown in **Figure 5**, some participants demonstrated almost no attention-related differences in alpha activity, while a handful displayed consistent and pronounced effects. To be sure, this inter-individual variability is not a new difficulty in the field of BCI. User-centered design has long been an important topic (Akçakaya et al., 2014; Moghadamfalahi et al., 2015), and even some of the few lateralized posterior alpha designs have

documented individual differences among users (Horschig et al., 2015). A viable approach to an alpha-compatible BCI likely includes identification of alpha “responders,” or individuals who demonstrate target-related changes in alpha activity, prior to including alpha activity as a classifiable signal in any machine learning model. To this point, a study from van Gerven and Jensen (2009) characterized different patterns of classification performance in participants sorted according to their “good” or “bad” alpha responses. Prior work has also demonstrated the benefits of personalizing channel selection to measure posterior alpha for BCI control (van Gerven et al., 2009).

Electroencephalography Ensemble Methods

With high-dimensional data such as EEG, one alternative to classification using a single model is to instead use ensemble learning to combine multiple EEG classification approaches (Sun et al., 2007). Unlike multi-modal approaches that integrate different input modalities (e.g., EEG, EOG, EMG) to improve BCI function (Li et al., 2016), ensemble learning can deploy multiple analytic models for a single input. To this end, the current RSVP speller might be altered to consider an ensemble that simultaneously includes both ERP information and posterior alpha activity. Indeed, recent research has combined event-related frequency band information with ERP information for ensemble classification in a motor imagery task (Luo et al., 2020). A past investigation has also proposed that posterior alpha may be simultaneously useful as a marker of both attentional engagement and abrupt changes in mental state, including frustration (Myrden and Chau, 2017). Given the

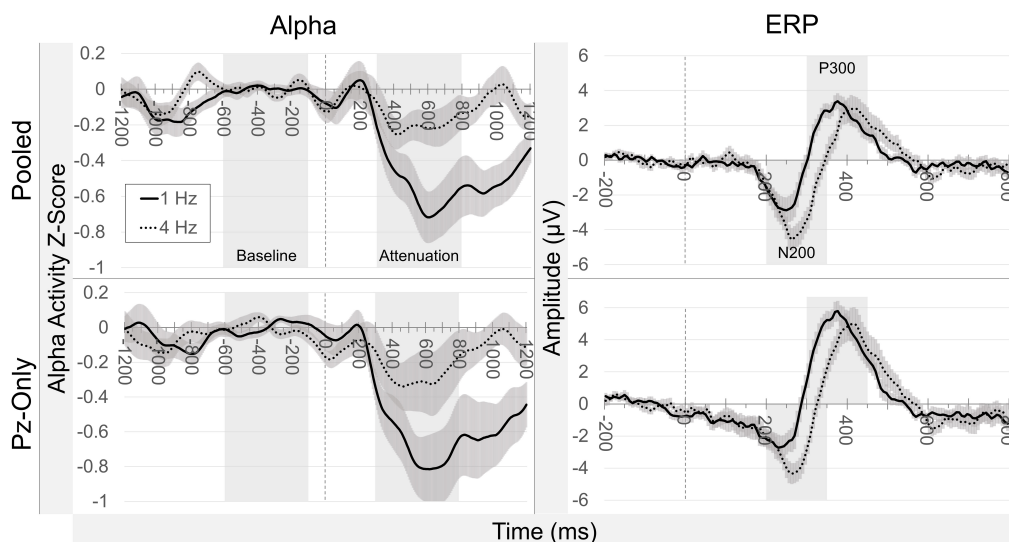


FIGURE 7 | Difference waveforms. Grand average difference waveforms illustrate target minus non-target signals. Alpha attenuation (column 1) and ERP target effects (column 2) are separated according to signal source and presentation rate condition. Highlighted temporal windows correspond to baselines and measured attenuation (alpha), and default peak detection settings (ERPs). There is a clear reduction in alpha attenuation in the 4 Hz condition, while N200 and P300 target effects remain relatively similar in magnitude between presentation rates.

comparatively low performance of the alpha classifiers in this study, it seems prudent to explore integration of alpha and ERPs to improve classification of user intent in RSVP and other comparable paradigms.

LIMITATIONS AND FUTURE DIRECTIONS

Generalizability of this study is hindered by a small sample size ($n = 12$). Participants were generally healthy adults, so it remains to be seen whether our results extend to individuals

with locked-in syndrome or other clinical populations. The use of pooled occipital sites for alpha and ERP analyses was relatively coarse, and future research would benefit from use of a recording apparatus of higher spatial resolution in order to evaluate topographic features of alpha attenuation in the context of RSVP. Artifact rejection methods resulted in only marginal changes to the data, possibly due to changes in statistical power after removing trials, since the across-participant measures did not change significantly following these rejections. Regardless, we recommend using these artifact rejection procedures consistently in future investigations. While overlapping processing of sequential visual stimuli is an inherent consequence of the RSVP design, the overlap was clearly detrimental to our measures of alpha activity in the 4 Hz condition. Additional work may benefit from attempts at noise reduction. A possible solution for this phenomenon would be to jitter stimulus duration to attenuate the SSVEP signal and associated harmonics. Lastly, though care was taken to match the preprocessing steps in both BrainVision and Python, there are minor differences inherent in the use of two different software packages. Future research would benefit from the use of a single, integrated toolbox.

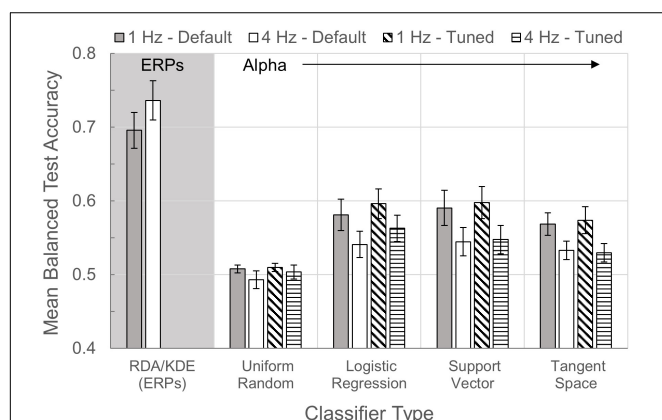


FIGURE 8 | Classifier comparisons. Bars illustrate the mean balanced test accuracies (\pm SEM) of multiple classifiers in the 1 and 4 Hz presentation conditions. With the exception of RDA/KDE, which quantifies performance of the ERP time-series classification, all other classifiers examine time-series alpha activity. The Uniform Random model was used as a control classifier with an expected classification accuracy of 0.50.

CONCLUSION

Event-related changes in posterior alpha activity are sensitive to visual attention during a common BCI speller paradigm, RSVP. Machine learning classifiers were able to discern target and non-target alpha responses at levels above chance, though these approaches were well below the current standards of more accepted ERP classifiers. However, the patterns of alpha

attenuation observed in this study were not redundant when compared to ERPs, even though these signals capture similar aspects of visual attention. These findings also offer data on posterior alpha in a paradigm that differs from previous investigations which rely on covert spatial attention. For these reasons, posterior alpha activity is a viable candidate signal for inclusion in future BCI designs, most expectedly in ensemble with other EEG control measures.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

This study involving human participants was reviewed and approved by the OHSU IRB. The participants provided their written informed consent to participate in the study.

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
- Akçakaya, M., Peters, B., Moghadamfalahi, M., Mooney, A. R., Orhan, U., Oken, B., et al. (2014). Noninvasive brain-computer interfaces for augmentative and alternative communication. *IEEE Rev. Biomed. Eng.* 7, 31–49. doi: 10.1109/RBME.2013.2295097
- Barachant, A., Bonnet, S., Congedo, M., and Jutten, C. (2013). Classification of covariance matrices using a Riemannian-based kernel for BCI applications. *Neurocomputing* 112, 172–178. doi: 10.1016/j.neucom.2012.12.039
- Barachant, A., Congedo, M., Jutten, C., and Brain, C. J. M. (2012). Multi-class brain computer interface classification by riemannian geometry. *IEEE Trans. Biomed. Eng.* 59, 920–928. doi: 10.1016/j.neucom.2017.02.014
- Bazanov, O. M., and Vernon, D. (2014). Interpreting EEG alpha activity. *Neurosci. Biobehav. Rev.* 44, 94–110. doi: 10.1016/j.neubiorev.2013.05.007
- Brain Products (2014). *BrainVision Analyzer. User Manual*. Gilching: Brain Products GmbH, 1–618.
- Cantero, J. L., Atienza, M., and Salas, R. M. (2002). Human alpha oscillations in wakefulness, drowsiness period, and REM sleep : different electroencephalographic phenomena within the alpha band. *Neurophysiol. Clin.* 32, 54–71. doi: 10.1016/s0987-7053(01)00289-1
- Corcoran, A. W., Alday, P. M., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2018). Toward a reliable, automated method of individual alpha frequency (IAF) quantification. *Psychophysiology* 55:e13064. doi: 10.1111/psyp.13064
- Foster, J. J., Sutterer, D. W., Serences, J. T., Vogel, E. K., and Awh, E. (2017). Alpha-band oscillations enable spatially and temporally resolved tracking of covert spatial attention. *Psychol. Sci.* 28, 929–941. doi: 10.1177/0956797617699167
- Fried-Oken, M., Kinsella, M., Peters, B., Eddy, B., and Wojciechowski, B. (2020). Human visual skills for brain-computer interface use: a tutorial. *Disabil. Rehabil. Assist. Technol.* 15, 799–809. doi: 10.1080/17483107.2020.1754929
- Grent-T-Jong, T., Boehler, C. N., Kenemans, J. L., and Woldorff, M. G. (2011). Differential functional roles of slow-wave and oscillatory-alpha activity in visual sensory cortex during anticipatory visual-spatial attention. *Cereb. Cortex* 21, 2204–2216. doi: 10.1093/cercor/bhq279
- Herscovitch, J., and Broughton, R. (1981). Sensitivity of the stanford sleepiness scale to the effects of cumulative partial sleep deprivation and recovery oversleeping. *Sleep* 4, 83–92. doi: 10.1093/sleep/4.1.83

AUTHOR CONTRIBUTIONS

BO, DK, and TM designed the study protocol. DK and BO planned the statistical analysis. DK collected data, processed EEG recordings, performed statistical tests, and wrote the first draft of the manuscript. DE, NS-M, and BC designed the classification analysis. NS-M wrote the primary Python analysis script. NS-M, BC, and TM consulted on the code. TM, NS-M, and BC wrote portions of the manuscript. All authors contributed to revising the manuscript and approved the final draft.

FUNDING

Support for this research was provided in part by NIH DC009834.

ACKNOWLEDGMENTS

Andy Fish, for administrative support; Deirdre Galvin-McLaughlin, for development of the RSVP practice task; CAMBI and all associated researchers.

- Horschig, J. M., Oosterheert, W., Oostenveld, R., and Jensen, O. (2015). Modulation of posterior alpha activity by spatial attention allows for controlling a continuous brain-computer interface. *Brain Topogr.* 28, 852–864. doi: 10.1007/s10548-014-0401-7
- Huang, Y., Erdogmus, D., Pavel, M., Mathan, S., and Hild, K. E. (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
- Kelly, S. P., Lalor, E., Reilly, R. B., and Foxe, J. J. (2005). “Independent brain computer interface control using visual spatial attention-dependent modulations of parieto-occipital alpha,” in *Proceedings of 2nd International IEEE EMBS Conference on Neural Engineering*, Vol. 2005 (Arlington, VA: IEEE), 667–670. doi: 10.1109/CNE.2005.1419713
- Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res. Rev.* 29, 169–195. doi: 10.1016/S0165-0173(98)00056-3
- Krusienski, D. J., Sellers, E. W., McFarland, D. J., Vaughan, T. M., and Wolpaw, J. R. (2008). Toward enhanced P300 speller performance. *J. Neurosci. Methods* 167, 15–21. doi: 10.1016/j.jneumeth.2007.07.017
- Lopes da Silva, F. (2013). EEG and MEG: relevance to neuroscience. *Neuron* 80, 1112–1128. doi: 10.1016/j.neuron.2013.10.017
- Lopes da Silva, F. H., Vos, J. E., Mooibroek, J., and van Rotterdam, A. (1980). Relative contributions of intracortical and thalamo-cortical processes in the generation of alpha rhythms, revealed by partial coherence analysis. *Electroencephalogr. Clin. Neurophysiol.* 50, 449–456. doi: 10.1016/0013-4694(80)90011-5
- Lee, G. R., Gommers, R., Waselewski, F., Wohlfahrt, K., and O’Leary, A. (2019). Pywavelets: a python package for wavelet analysis. *J. Open Source Softw.* 4:1237. doi: 10.21105/joss.01237
- 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, 1–23. doi: 10.1088/1741-2552/aa9817
- Li, Y., Pan, J., Long, J., Yu, T., Wang, F., Yu, Z., et al. (2016). Multimodal BCIs: target detection, multidimensional control, and awareness evaluation in patients with disorder of consciousness. *Proc. IEEE* 104, 332–352. doi: 10.1109/JPROC.2015.2469106
- Luo, J., Gao, X., Zhu, X., Wang, B., Lu, N., and Wang, J. (2020). Motor imagery EEG classification based on ensemble support vector learning. *Comput. Methods Programs Biomed.* 193:105464. doi: 10.1016/j.cmpb.2020.105464

- MacLean, A. W., Fekken, G. C., Saskin, P., and Knowles, J. B. (1992). Psychometric evaluation of the stanford sleepiness scale. *J. Sleep Res.* 1, 35–39. doi: 10.1111/j.1365-2869.1992.tb00006.x
- Meirovitch, Y., Harris, H., Dayan, E., Arieli, A., and Flash, T. (2015). Alpha and beta band event-related desynchronization reflects kinematic regularities. *J. Neurosci.* 35, 1627–1637. doi: 10.1523/JNEUROSCI.5371-13.2015
- Memmott, T., Koçanaoğulları, A., Lawhead, M., Klee, D., Dudy, S., Fried-Oken, M., et al. (2021). BciPy: brain-computer interface software in python. *Brain Comput. Interfaces* 8, 1–18. doi: 10.1080/2326263X.2021.1878727
- Moghadamfalahi, M., Orhan, U., Akcakaya, M., Nezamfar, H., Fried-Oken, M., and Erdogmus, D. (2015). Language-model assisted brain computer interface for typing : a comparison of matrix and rapid serial visual presentation. *IEEE Trans. Neural Syst. Rehabil. Eng.* 23, 910–920. doi: 10.1109/TNSRE.2015.2411574
- Myrden, A., and Chau, T. (2017). A passive EEG-BCI for single-trial detection of changes in mental state. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 345–356. doi: 10.1109/TNSRE.2016.2641956
- Okazaki, Y. O., Horschig, J. M., Luther, L., Oostenveld, R., Murakami, I., and Jensen, O. (2015). Real-time MEG neurofeedback training of posterior alpha activity modulates subsequent visual detection performance. *NeuroImage* 107, 323–332. doi: 10.1016/j.neuroimage.2014.12.014
- Oken, B., Memmott, T., Eddy, B., Wiedrick, J., and Fried-Oken, M. (2018). Vigilance state fluctuations and performance using brain-computer interface for communication. *Brain Comput. Interfaces* 5, 146–156. doi: 10.1080/2326263X.2019.1571356
- Oken, B. S., Orhan, U., Roark, B., Erdogmus, D., Fowler, A., Mooney, A., et al. (2014). Brain-computer interface with language model-electroencephalography fusion for locked-in syndrome. *Neurorehabil. Neural Repair* 28, 387–394. doi: 10.1177/1545968313516867
- Orhan, U., Erdogmus, D., Roark, B., Oken, B., and Fried-oken, M. (2013). Offline analysis of context contribution to ERP-based typing BCI performance. *J. Neural Eng.* 10:066003. doi: 10.1088/1741-2560/10/6/066003
- Orhan, U., Erdogmus, D., Roark, B., Purwar, S., Hild Ii, K. E., Oken, B., et al. (2011). “Fusion with language models improves spelling accuracy for erp-based brain computer interface spellers,” in *Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Boston, MA: IEEE), 5774–5777. doi: 10.1109/IEMBS.2011.6091429
- Orhan, U., Hild, K. E., Erdogmus, D., Roark, B., Oken, B., and Fried-Oken, M. (2012). RSVP keyboard: an EEG based typing interface. *Proc. IEEE Int. Conf. Acoust Speech Signal Process* 645–648. doi: 10.1109/ICASSP.2012.6287966
- Patel, S. H., and Azzam, P. N. (2005). Characterization of N200 and P300: selected studies of the event-related potential. *Int. J. Med. Sci.* 2, 147–154. doi: 10.7150/ijms.2.147
- Peirce, J. W. (2007). Psychopy-psychophysics software in python. *J. Neurosci. Methods* 162, 8–13. doi: 10.1016/j.jneumeth.2006.11.017
- Peirce, J. W. (2009). Generating stimuli for neuroscience using psychopy. *Front. Neuroinform.* 2, 1–8. doi: 10.3389/neuro.11.010.2008
- Peng, W., Hu, L., Zhang, Z., and Hu, Y. (2012). Causality in the association between P300 and alpha event-related desynchronization. *PLoS One* 7:e34163. doi: 10.1371/journal.pone.0034163
- Pfurtscheller, G., Brunner, C., Schlögl, A., and Lopes da Silva, F. H. (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., 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
- Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019
- Rashid, M., Sulaiman, N., P. P. Abdul Majeed, A., Musa, R. M., Ahmad, A. F., Bari, B. S., et al. (2020). Current status, challenges, and possible solutions of eeg-based brain-computer interface: a comprehensive review. *Front. Neurobot.* 14:25. doi: 10.3389/fnbot.2020.00025
- Rezeika, A., Benda, M., Stawicki, P., Gembler, F., Saboor, A., and Volosyak, I. (2018). Brain-computer interface spellers: a review. *Brain Sci.* 8:57. doi: 10.3390/brainsci8040057
- Sun, S., Zhang, C., and Zhang, D. (2007). An experimental evaluation of ensemble methods for EEG signal classification. *Pattern Recognit. Lett.* 28, 2157–2163. doi: 10.1016/j.patrec.2007.06.018
- Treder, M. S., Bahramisharif, A., Schmidt, N. M., Van Gerven, M. A., and Blankertz, B. (2011). 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
- van Gerven, M., Bahramisharif, A., Heskes, T., and Jensen, O. (2009). Selecting features for BCI control based on a covert spatial attention paradigm. *Neural Networks* 22, 1271–1277. doi: 10.1016/j.neunet.2009.06.004
- 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
- Vázquez-Marrufo, M., García-Valdecasas, M., Caballero-Díaz, R., Martín-Clemente, R., and Galvao-Carmona, A. (2019). Multiple evoked and induced alpha modulations in a visual attention task: latency, amplitude and topographical profiles. *PLoS One* 14:e0223055. doi: 10.1371/journal.pone.0223055
- Wolpaw, J. R. (2013). “Brain-computer interfaces,” in *Handbook of Clinical Neurology*, 1st Edn, Vol. 110 (Amsterdam: Elsevier), 15–23. doi: 10.1016/B978-0-444-52901-5.00006-X
- Wolpaw, J. R., and Wolpaw, E. W. (2012). in *Brain-Computer Interfaces: Principles and Practice*, eds J. R. Wolpaw and E. W. Wolpaw (Oxford: Oxford University Press). doi: 10.1093/acprof:oso/9780195388855.001.0001
- Yordanova, J., Kolev, V., and Polich, J. (2001). P300 and alpha event-related desynchronization (ERD). *Psychophysiology* 38, 143–152. doi: 10.1017/S0048577201990079
- Zenodo (2022). CAMBI-Tech/Alpha-Attenuation: Initial Release (1.0.0). Geneva: CERN. doi: 10.5281/zenodo.6098824

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.

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.

Copyright © 2022 Klee, Memmott, Smedemark-Margulies, Celik, Erdogmus and Oken. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Brain–Heart Interaction and the Experience of Flow While Playing a Video Game

Shiva Khoshnoud^{1,2*}, Federico Alvarez Igarzábal¹ and Marc Wittmann¹

¹ Institute for Frontier Areas of Psychology and Mental Health, Freiburg, Germany, ² Department of Neurosurgery and Neurotechnology, Institute for Neuromodulation and Neurotechnology, University of Tübingen, Tübingen, Germany

OPEN ACCESS

Edited by:

Filippo Zappasodi,
University of Studies G. d'Annunzio
Chieti and Pescara, Italy

Reviewed by:

Enrica Laura Santarcangelo,
University of Pisa, Italy
Almira M. Kustubayeva,
Al-Farabi Kazakh National University,
Kazakhstan
Maurizio Bertollo,
G. d'Annunzio University
of Chieti–Pescara, Italy

*Correspondence:

Shiva Khoshnoud
khoshnoud@igpp.de

Specialty section:

This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

Received: 22 November 2021

Accepted: 06 April 2022

Published: 28 April 2022

Citation:

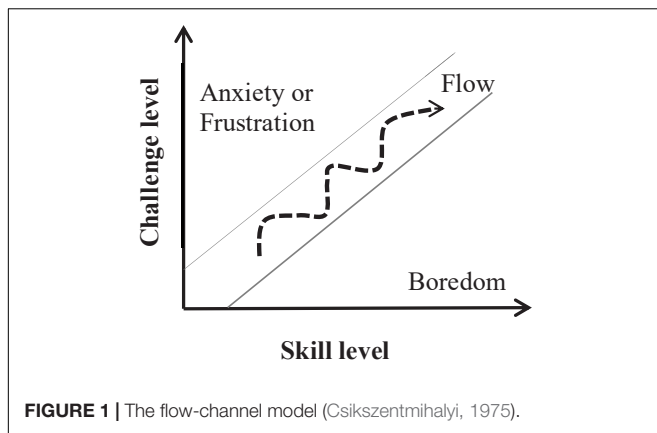
Khoshnoud S, Alvarez Igarzábal F
and Wittmann M (2022) Brain–Heart
Interaction and the Experience
of Flow While Playing a Video Game.
Front. Hum. Neurosci. 16:819834.
doi: 10.3389/fnhum.2022.819834

The flow state – an experience of complete absorption in an activity – is linked with less self-referential processing and increased arousal. We used the heart-evoked potential (HEP), an index representing brain–heart interaction, as well as indices of peripheral physiology to assess the state of flow in individuals playing a video game. 22 gamers and 21 non-gamers played the video game Thumper for 25 min while their brain and cardiorespiratory signals were simultaneously recorded. The more participants were absorbed in the game, the less they thought about time and the faster time passed subjectively. On the cortical level, the fronto-central HEP amplitude was significantly lower while playing the game compared to resting states before and after the game, reflecting less self-referential processing while playing. This HEP effect corresponded with lower activity during gameplay in brain regions contributing to interoceptive processing. The HEP amplitude predicted the level of absorption in the game. While the HEP amplitude was overall lower during the gaming session than during the resting states, within the gaming session the amplitude of HEP was positively associated with absorption. Since higher absorption was related to higher performance in the game, the higher HEP in more absorbed individuals reflects more efficient brain–heart interaction, which is necessary for efficient game play. On the physiological level, a higher level of flow was associated with increased overall sympathetic activity and less inhibited parasympathetic activity toward the end of the game. These results are building blocks for future neurophysiological assessments of flow.

Keywords: experience of flow, heart-evoked potential, brain–heart interaction, self-referential processing, arousal, time perception

INTRODUCTION

Csikszentmihalyi (1975) described the state of flow as an optimal experience that occurs when an individual is immersed in an activity and feels enjoyment. The main antecedents of this experiential state are clear goals, immediate feedback, and an optimal balance between the individual's skills and the level of challenge posed by the activity (the skill–challenge balance; Csikszentmihalyi and Csikszentmihalyi, 1992; Engeser and Rheinberg, 2008; Keller and Blomann, 2008; Fong et al., 2015). According to the flow-channel model (Csikszentmihalyi, 1975, 1990; **Figure 1**), whenever the challenge level of the activity outweighs the performer's skill level, the person will become frustrated or anxious. In contrast, when the level of challenge is lower than the individual's skill level, they will feel bored. The flow experience has been associated with successful performance and feelings of competence (Engeser and Rheinberg, 2008; Jin, 2012; Rutrecht et al., 2021),



since frustration and boredom lead to diminished concentration and, consequently, poor performance (Perone et al., 2019). Flow is often linked with the experience of losing the sense of time and of the self (Csikszentmihalyi and Csikszentmihalyi, 1992; Wittmann, 2015, 2018; Rutrecht et al., 2021).

The underlying physiological and neural mechanisms of the flow state are beginning to be studied extensively. Researchers have attempted to map specific neurophysiological indicators of this mental state by employing a skill-challenge balance. So far, the results have been heterogeneous and inconclusive in terms of identifying a unified mechanism underlying this state (for more details see our review, Khoshnoud et al., 2020). In the present study, we focused on two important mechanisms of the experience of flow, namely heightened arousal as manifested in elevated cardiovascular (de Manzano et al., 2010; Chanel et al., 2011; Keller et al., 2011; Tian et al., 2017; De Sampaio Barros et al., 2018) and loss of self-related processing typically associated with decreased activity in the default-mode network (DMN, Ulrich et al., 2014, 2016, 2018; De Sampaio Barros et al., 2018).

Video games are promising tools for systematically inducing flow under controlled laboratory conditions. They offer clear goals, immediate feedback, and challenging tasks with the possibility of modulating the level of difficulty for achieving the desired skill-challenge balance (Salen and Zimmerman, 2003; Alvarez Igarzábal, 2019). For this reason, we chose to investigate arousal and self-related processing during flow states with the commercially available video game *Thumper*, which our previous research had shown to be highly flow inducing (Rutrecht et al., 2021). We tested experienced gamers and non-gamers who played the game for 25 min and recorded their electroencephalography (EEG) and cardiorespiratory signals. Gamers were recruited because their experience with video games would in principle allow them to more easily familiarize themselves with the game and enter a flow state. The non-gamer sample constituted the control group to allow us to compare how individuals lacking gaming skills would experience the play sessions – i.e., if they would enter a state of flow or not. This study was conducted in the context of the VIRTUALTIMES project with the goal of developing a therapeutic game that could reduce the symptoms of patients with psychiatric disorders through the induction of flow states. Recent conceptualizations suggest

that inducing the experience of flow with video games could help reduce pathologically increased self-rumination (Kühn et al., 2018), which is associated with time distortions in many patient groups with psychopathologies, such as depression, attention deficit hyperactivity disorder, and addiction (Wittmann et al., 2007; Khoshnoud et al., 2018; Vogel et al., 2018). The game resulting from the VIRTUALTIMES project should effectively induce flow states regardless of the gaming experience of the player. We included the non-gamer sample for this reason as well, since potential differences between groups (or a lack thereof) would inform the design of the game. Continuous play, which is the more natural way of playing video games – by progressing and encountering increasingly difficult challenges – was preferred over difficulty modulation, in which the level of challenge is adapted to the player's performance. Employing such form of difficulty modulation to achieve the skill-challenge balance can facilitate flow induction, but does not necessarily lead to a state of flow.

Flow, Selflessness, and Brain–Heart Interaction

Reduced self-awareness is one of the main characteristics of the flow experience (Csikszentmihalyi, 1975, 1990). The high levels of concentration and focused attention demanded by the task at hand restrict resource allocation for task-irrelevant demands, like body- and self-awareness. Several studies reported less self-referential processing during the flow experience by showing deactivation of the DMN, specifically the medial prefrontal cortex (MPFC) (Sadlo, 2016; Ulrich et al., 2016, 2014, 2018; De Sampaio Barros et al., 2018; Ju and Wallraven, 2019). DMN activity is associated with relaxation, mind-wandering, and self-referential thinking, and it diminishes during task-focused and goal-directed actions (Raichle et al., 2001; Goldberg et al., 2006). Increased activation of the insular cortex has also been reported during flow states (Ulrich et al., 2016; Huskey et al., 2018; Ju and Wallraven, 2019). The insular cortex is the primary visceral area (Craig, 2009). Recent neuroscientific research has shown that cortical processing of signals from internal visceral organs, like the heart and the gut, are important for cognition (Park et al., 2014; Critchley and Garfinkel, 2018), self-consciousness (Damasio and Carvalho, 2013; Tallon-Baudry et al., 2018; Park and Blanke, 2019), and subjective time (Craig, 2009; Wittmann, 2013; Teghil et al., 2020). According to Tallon-Baudry et al. (2018), the gut and the heart are similar to ticking clocks that constantly send intrinsically generated information up to the central nervous system. By monitoring this bodily information, the brain creates a neural reference frame for developing a first-person perspective.

To evaluate self-related processing during moments of flow, we investigated the brain–heart interaction by exploring possible associations between the heart-evoked potential (HEP) – an index representing the neural processing of cardiac afferents – and the flow experience while playing the game. HEPs are cortical electrophysiological responses in the brain that are time-locked to the R-peaks of the simultaneously measured electrocardiography (ECG) signal (Montoya et al., 1993; Schandry and Montoya, 1996). The precise pathways underlying

the HEP are unknown. In addition to cardiac and blood vessel receptors, it has been argued that cardiac afferents are projected to the cortex through tactile and proprioceptive receptors (Azzalini et al., 2019; Park and Blanke, 2019). Research has shown that self-relatedness (Babo-Rebelo et al., 2016, 2019; Park et al., 2016; Sel et al., 2017), focus of attention (Petzschnner et al., 2019), emotional feelings (Fukushima et al., 2011; Shao et al., 2011), and arousal (Luft and Bhattacharya, 2015) modulate the HEP amplitude.

Despite the established relationship between selfhood and the HEP amplitude, the findings regarding the direction of this association are mixed; i.e., there are positive and negative correlations between the HEP amplitude and self-relatedness. Employing a full-body illusion paradigm, Park et al. (2016) identified a relationship between the HEP amplitude over fronto-central scalp sensors and experimentally modulated changes in self-identification. The negative HEP amplitude was more pronounced during the condition with lower self-identification rating compared to the condition with higher self-identification. A study conducted by Babo-Rebelo et al. (2016) revealed a direct link between self-relatedness of spontaneous thoughts and the HEP amplitude. The HEP amplitude was evaluated in two conditions: one where participants were the subject (first-person perspective) of presented thoughts (“I” as in “I like him”) and another where they were the object of the thoughts (“Me” as in “He likes me”). The amplitude of the HEP in the mid-posterior regions of the brain co-varied negatively with the engagement of “I”, showing higher amplitudes for lower ratings on the “I” scale. The HEP amplitude also differed between “high” and “low” trials on the “Me” scale over medial frontal sensors with higher HEP amplitudes for the trials with higher score of the “Me” scale.

In the present work, we investigated the link between flow and the HEP amplitude as an index reflecting self-related processing. This was the first time that the HEP was studied in the context of flow, as far as we know. Considering the connection between the HEP amplitude and the self (Park et al., 2016; Babo-Rebelo et al., 2019, 2016), as well as the existing overlap between the cortical sources of the HEP and the DMN (Park and Blanke, 2019), one can expect to see a decrease in neural responses to the heartbeats and a lower HEP amplitude while playing a video game compared to the resting state. This means that self-referential processing is lower while gaming than while resting (*Hypothesis 1*). The aforementioned associations between selfhood and the amplitude of the HEP may refer to a distinct HEP during gameplay as a result of experiencing flow and a loss of self-awareness. Considering the mixed nature of the direction of the reported associations, we expected a relationship between flow and the HEP amplitude without making any assumptions regarding its direction (*Hypothesis 2*).

Flow and Arousal

The feeling of enjoyment along with high levels of concentration on a given task suggests that flow states are modulated through arousal levels, which are related to the activation of the autonomic nervous system (ANS). The two branches of the ANS, the sympathetic nervous system (SNS, responsible for “fight-or-flight” responses) and the parasympathetic nervous system

(PNS, responsible for “rest-and-digest” responses) operate as excitatory and inhibitory physiological mechanisms, respectively (Cacioppo et al., 2007). Numerous indicators of sympathetic and parasympathetic activity have been studied in flow research, including cardiovascular, electrodermal, and respiratory measures. Aggregating the existing empirical results, both linear and non-linear associations between the activity of the autonomic nervous system and flow have been reported (Khoshnoud et al., 2020). The most established theory links the experience of flow to high levels of the SNS activity (Kivikangas, 2006; de Manzano et al., 2010). The reported positive association between the electrodermal activity (EDA) as a robust indicator of sympathetic arousal (Critchley and Nagai, 2013) and flow (Nacke and Lindley, 2010; Léger et al., 2014; Ulrich et al., 2016, 2014) confirms this theory. Increased flow while playing a video game was reported to be related to an increased heart rate (de Manzano et al., 2010; Bian et al., 2016), a faster respiration rate (Bian et al., 2016; Tian et al., 2017), high levels of salivary cortisol (Keller et al., 2011), and lower heart-rate variability (HRV) measures, specifically lower high-frequency (HF) HRV (Chanel et al., 2011; Keller et al., 2011; Harmat et al., 2015; Harris et al., 2016; De Sampaio Barros et al., 2018; Kozhevnikov et al., 2018). HF-HRV reflects the variance in the 0.15–0.4 Hz frequency range of the heart rate and is a reliable indicator of parasympathetic activity (Laborde et al., 2017; Shaffer and Ginsberg, 2017).

In two studies, the reported heightened sympathetic activity during flow was associated with the modulated parasympathetic activity manifested by deeper respiration (de Manzano et al., 2010; Harmat et al., 2015; Tian et al., 2017). According to Porges’ (1995, 2021) polyvagal theory, parasympathetic influences are essential for an individual’s successful adaptation to changing environmental demands. Several studies have suggested a connection between parasympathetic activity and cognitive performance, working memory, and attention (Hansen et al., 2003; Frewen et al., 2013; Mahinrad et al., 2016; Tsakiris and De Preester, 2018). The so-called parasympathetic modulation of sympathetic activity may describe flow as a state of heightened arousal accompanied by a feeling of pleasure and, at the same time, seemingly paradoxical relaxation, which is suggested to act as a physiological coping mechanism for high demands of attention in moments of flow (de Manzano et al., 2010; Ullén et al., 2010).

In light of the above-mentioned findings, we expected to see a positive association between sympathetic activity and the self-reported flow measures. Given the reported connection between parasympathetic activity and attention, relatively increased parasympathetic activity should help cope with the changing attentional demands of the task. We assumed that more flow would be associated with increased sympathetic and parasympathetic activity (*Hypothesis 3*).

MATERIALS AND METHODS

Participants

Participants were 43 healthy subjects (Gamers: $n = 22$; 19 males, 3 females; Non-Gamers: $n = 21$; 18 males, 3 females) with

an average age of 24.90 ± 3.98 years (Gamers: 25.59 ± 4.56 , Non-Gamers: 24.19 ± 3.23). The age difference between the two groups was not significant ($t(41) = -1.15$, $p = 0.254$). All participants were recruited via online platforms (Student Services University Freiburg), flyers, word-of-mouth dissemination, and through advertisements on social media and gaming forums. As an inclusion criterion for being a gamer, individuals had to: (a) consider themselves a gamer, (b) have more than five years of experience playing videogames regularly, and (c) have played videogames in the past six months for at least 5 h a week. On average, gamers had 15.36 ± 4.97 years of experience and non-gamers, 6.28 ± 7.3 years with an average of 15.97 ± 8.89 and 0.48 ± 0.54 h of playing games in a week, respectively. Gaming experience in years of playing ($t(41) = -4.77$, $p < 0.001$) and average hours of gameplay in a week ($t(40) = -7.96$, $p < 0.001$) were significantly higher in gamers than in non-gamers. No significant difference was identified for the educational level between the two groups ($U = 263$, $p = 0.4$). This study was approved by the local Ethics Committee of the Institute for Frontier Areas of Psychology and Mental Health (IGPP_2019_01).

Questionnaires

The state of flow was measured with the 10-item *Flow Short Scale* (FSS, Rheinberg and Vollmeyer, 2003) with seven-point Likert scales ranging from 1 (not at all) to 7 (very much) which are linked to two subscales: absorption and fluency. *Absorption* is measured by four items (e.g., “I am totally absorbed in what I am doing” or “I don’t notice time passing”), while the remaining six items index the *fluency* of performance (e.g., “My thoughts/activities run fluidly and smoothly” or “The right thoughts/movements occur of their own accord”). Considering the flow-channel model (Figure 1), the absorption subscale can be considered a stronger indicator of flow, while fluency can also occur in situations in which external demands are lower than individual skills (Peifer et al., 2014). Besides the overall flow score comprising all items (FSS mean score), we also looked at the absorption and fluency subscales’ mean values separately in their relationships with the other behavioral and physiological measures. The Cronbach’s alpha values found in our dataset for the FSS mean score, the absorption, and the fluency subscales are 0.847, 0.768, and 0.787, respectively.

We used the *Subjective Time, Self, and Space* (STSS) questionnaire (Jokic et al., 2018) to assess the effects of the gaming session on time and bodily perception. This questionnaire consists of an item to evaluate bodily awareness (“How intensively did you experience your body most of the time?”) with a non-verbal pictorial scale with the answer category ranging from 1 to 7. Three items assess the experience of time: (1) “Intuitively, without thinking about it, the gameplay session lasted ____ minutes and ____ seconds” (estimated duration); (2) “How often did you think about time?” (thinking about time); and (3) “How fast did time pass?” (speed of time passage). The latter two time questions were answered with a vertical streak on a visual analog scale (a 10-centimeter-long horizontal line) ranging from “not at all” to “a lot” on question two and from “very slowly”

to “very fast” on question three. The *Self-Assessment Manikin* (SAM) scale (Bradley and Lang, 1994) was used to measure *valence* (SAM-valence) and *arousal* (SAM-arousal) changes with non-verbal, pictorial scales after vs. before the gaming session. We used the 5-point version of the scales.

Experimental Design

The study took place on 3 consecutive days. The first two days were training sessions for participants to learn how to play the chosen game and on the third day the participants’ EEG and cardiorespiratory signals were recorded simultaneously while they played the game. The video game we used was *Thumper*, an action/rhythm game developed by the studio Drool and released in 2016. We acquired and launched the game through the platform Steam. In this game, the player controls a silver beetle that moves forward automatically on a track within an abstract landscape. The player watches from behind the beetle, so that s/he can react to different elements that are placed on the track. The track also presents the player with obstacles like sharp curves or spikes that can damage the beetle if they are not averted in time. Players can listen to the rhythm induced by the soundtrack to help with the timing of these actions. The goal of this game is to achieve the highest score possible in each level and level section by hitting as many lights as possible and avoiding collisions with obstacles.

The first session lasted 90 min in which participants were asked to read the information sheet, sign the consent form, and answer questions regarding their well-being. After that the experimenter explained the gameplay basics to the participants and let them play for about 60 min, starting from the first level of the game. In the second training session, participants played the game for another 60 min, starting from the level they had reached in the previous session, unless they preferred to start at a lower level. In the final recording session, participants could choose the level at which they felt comfortable to start. We recommended that they neither pick a level that was too challenging nor one that was too easy for them. This was done to increase the chances that they would enter a flow state during the gameplay session. Gamers and non-gamers therefore played at different game levels according to their skills, since higher levels are more challenging than lower ones.

In the final session, participants were asked to sit in a comfortable chair watching a screen positioned 70 cm in front of them in an electrically shielded room. They were requested to fill out the SAM questionnaire after all the electrodes had been placed. The recording session had three stages: a pre-game resting-state period of 5 min with eyes open, followed by 25 min of gameplay, and then another 5 min with a post-game, resting-state recording. Participants were asked to play *Thumper* continuously for 25 min, since there seems to be an adaptation period while playing after which participants tend to lose track of time (Tobin et al., 2010; Bisson et al., 2012). Yun et al. (2017) reported that 87.5% of the participants in their study required at least 25 min to get into the flow state. After the post-game session, participants filled out the SAM questionnaire once again, as well as the post-task questionnaires STSS, and FSS. After the task, they watched their

own recorded gameplay and answered questions regarding their subjective experience.

Signal Recordings

Continuous EEG signals were recorded using a 32-channel electrocap (ActiCHamp, BrainVision) with active electrodes positioned according to the extended 10–10 international system. All electrodes were referenced to the Fz electrode, with the ground electrode placed on the forehead. Electrode impedances were kept below 10 k Ω . EEG signals were digitized with a 1000 Hz sampling rate and band-pass filtered within the 0.01–120 Hz range. One electrocardiogram (ECG) signal was acquired using three Ag/AgCl electrodes, which were positioned according to the Lead II Einthoven configuration: two electrodes placed on the right clavicle and the left hip/abdomen (active electrodes), and one electrode placed on the left clavicle (ground electrode). A respiratory signal was acquired using the Brain Vision respiration belt attached to the participant's chest or abdomen, depending on the subject's breathing mode. All peripheral signals were co-registered with the EEG via the auxiliary inputs of the amplifier.

Data Processing

Electroencephalography Analysis

The recorded EEG signals were processed with the Matlab software using custom-written scripts and the EEGLAB toolbox functions. The raw EEG signals were first down-sampled to 250 Hz and filtered with a band pass filter of 1.5–70 Hz. After this initial filtering step, line noise and other large non-stationary artifacts were identified and cleaned using the artifact subspace reconstruction (ASR) approach (Chang et al., 2020). The cleaned signals were re-referenced to a common average reference. To identify other non-brain related EEG contamination (e.g., eye-blinks, muscle, heart, and channel noise), we used the adaptive mixture independent component analysis (AMICA; Palmer et al., 2006, 2011, 2008) to decompose the EEG signal into its independent components (ICs). AMICA has shown superior performance among the blind-source separation algorithms for EEG decomposition (Delorme et al., 2012). After decomposing the signals into ICs, their equivalent current dipoles were also computed using the three-shell, boundary-element-method head model based on the MNI brain template using the DIPFIT plugin of the EEGLAB toolbox¹. The identified ICs were then automatically classified and labeled using a machine-learning approach, which has been trained to classify the ICs based on several characteristics, such as spectral properties and brain topography (Pion-Tonachini et al., 2019). The brain-related ICs, which were labeled with the probability higher than 0.5 and had a dipole residual variance less than 0.2, were selected for further analysis. This procedure cleans the signal from eye-blinks, muscle noise, heart artifact, and other contamination and preserves pure brain-related activity. To calculate the average alpha frequency band (8 – 12 Hz) power, we used the Welch's power-spectral density estimation method (using the hamming window with the size of 2000 samples and 10% overlap between windows).

¹http://scn.ucsd.edu/wiki/A08:_DIPFIT

Electrocardiography Analysis

The raw ECG signals were imported to the Kubios Heart Rate Analysis Software (Kubios, Inc., University of Western Finland, Finland) to calculate inter-beat intervals (IBIs) of successive heartbeats and associated heart-rate variability measures. Recordings were first screened manually for ectopic and missing beats, after which the appropriate artifact-removal threshold available in Kubios was applied. The average IBI was calculated for each condition and each subject. Power-spectrum analysis was performed using the Fast Fourier Transformation (FFT) method provided by the software, and then the log power of the low-frequency HRV (LF-HRV; 0.04–0.15 Hz) and the high-frequency HRV (HF-HRV; 0.15–0.4 Hz) were computed. Since the HF-HRV is influenced by breathing rates below 9 or above 24 cycles per minute (Malik et al., 1996; Laborde et al., 2017), we used the obtained respiration rate from the respiratory signal to control for respiratory rates in these ranges. Participants whose breathing rates exceeded these limits were excluded from the frequency-domain HRV analysis.

Respiration Analysis

The recorded respiratory signals were first down sampled to 25 Hz and lowpass filtered using an infinite-impulse-response (IIR) lowpass filter with the order of 8 and passband frequency of 2 Hz. Since the respiratory signals were contaminated with large artifacts due to movements, the signal amplitudes outside the chosen threshold (mean \pm 2 \times S.D.) were replaced with the interpolated values. Then the local maxima were identified and the average respiration rate (RR-mean) and the standard deviation of the respiration rate (RR-STD) were calculated.

Heartbeat-Evoked Potentials

For the HEP analysis, cardiac R peaks of the ECG signal were detected offline and used as triggers for the EEG segmentation of the pre-processed EEG signals. The R-peak detection was performed by decomposing the ECG signals using discrete wavelet transform analysis with the “sym4” wavelets. The “sym4” wavelet resembles the QRS complex of ECG signals, which makes it a good choice for the QRS detection. Wavelet decomposition was performed for five levels, and then the signal approximation was built from the wavelet coefficients of levels 4 and 5. The exact location of R peaks was identified using the squared absolute values of the signal approximation and a peak-finding algorithm. EEG epochs were then extracted time-locked from 200 ms before to 600 ms after the detected R-peaks. Baseline correction was performed from –200 to –100 ms time-locked to the R peak for each epoch. The HEP signals were then averaged for each electrode, each condition, and each subject.

Source Localization

Cortical source reconstruction and surface visualization were implemented by the BrainStorm toolbox (version January 04, 2021; Tadel et al., 2011) on the Matlab software. After co-registration of the EEG sensors and the template anatomy (MNI brain template, ICBM152) for each participant, the forward model was calculated using the boundary element method from the open-source software OpenMEEG (Gramfort et al., 2010) on

the cortical surface of the template brain. Cortical currents for each subject and each condition were estimated by a distributed model consisting of 15,002 current dipoles using the weighted minimum-norm, current-estimation method with the dynamical statistical parametric mapping (dSPM) providing z-score cortical currents. The orientation of dipoles was considered constraint and normal to the cortex. The cortical currents were then spatially smoothed (7 mm) and averaged over 400 to 500 ms after the R-peak (in which a significant difference between the pre-game and the game was found). The anatomical description is based on the Desikan–Killiany (Desikan et al., 2006) and the Brodmann parcellations available in the BrainStorm toolbox.

Statistical Analysis

(1) For between-subject behavioral comparisons, we computed the independent *t*-test and, in non-normal distributions, the Mann–Whitney *U*-test. (2) Peripheral measure comparisons were performed using repeated-measure ANOVAs with condition (pre-game, game, and post-game) as a within-subject factor, group (gamers vs. non-gamers) as a between-subject factor, and each measure as a dependent variable. Degrees of freedom in repeated-measure ANOVAs were corrected according to the Greenhouse–Geisser method for violation of the sphericity assumption when appropriate. (3) Correlations were analyzed with Pearson's correlation coefficients (*r*) affording parametric assumptions, such as a normal distribution. Whenever we found a significant effect using the Shapiro–Wilk test for non-normality in one of the two variables, we reported the Spearman correlation coefficient (ρ). Analysis significance levels were set to $p < 0.05$ for each correlation. The false-discovery rate (FDR) method, a multiple-comparisons-correction procedure by Benjamini and Hochberg (1995) was used to control for multiple tests for each of the correlation tables.

(4) Differences in the HEPs between conditions and the corresponding cortical sources were tested using two statistical methods. One was the cluster-based permutation *t*-test implemented in the Fieldtrip toolbox, which is available in the BrainStorm toolbox. With this procedure, individual samples with a *t* value higher than threshold ($p < 0.05$, two tailed), are clustered in connected sets based on temporal and spatial adjacency. For each cluster, cluster-level statistics are assigned by taking the sum of the *t* values within a cluster, and then the maximum of the cluster-level statistics is selected for the evaluation of the null hypothesis. After shuffling the condition labels 10,000 times, the two-tailed Monte-Carlo *p* value corresponds to the proportion of elements in the distribution of shuffled, maximal cluster-level statistics that exceeds the observed maximum or minimum original cluster-level test statistics. As this method uses maximal cluster-level test statistics, it intrinsically controls for multiple comparisons in time and space (Park et al., 2014). Without *a priori* selection of a time window and region of interest, getting a significant result from this method does not establish significance of effect for latency or location (Sassenhagen and Draschkow, 2019). To test the differences of the HEP amplitude in the sensory domain for which we did not have an *a priori* assumption of a time window and scalp locations, we used a non-parametric permutation test

with an FDR correction for multiple comparisons. After finding the time window in which the HEP amplitude was significantly different between conditions, the cortical-source differences were tested with the cluster-based permutation *t* test and then the non-parametric permutation test.

RESULTS

Subjective and Behavioral Results

Gamers vs. Non-gamers

By taking into account the outcomes of the two prior training sessions, gamers started playing at significantly higher levels of the game ($t(41) = -4.342, p < 0.001$) and achieved significantly higher levels ($t(41) = -4.030, p < 0.001$) than non-gamers. No significant differences were observed in performance between the two groups in terms of total final score ($t(41) = -1.449, p = 0.155$) and total final error ($t(41) = 0.315, p = 0.754$), which accumulated during playing on the different levels of the game. As shown in **Supplementary Table 1** in the **Supplementary Data**, playing *Thumper* elicited comparably high levels of flow (FSS mean score) in both groups with the value of 4.86 ± 1.03 in non-gamers and 5.34 ± 0.77 in gamers (7 is the maximum value). There were no significant differences between the two groups in self-reported flow ($t(41) = -1.741, p = 0.089$) or for the subscales of absorption ($t(41) = -1.819, p = 0.076$) and fluency ($t(41) = -1.324, p = 0.193$). The estimated duration of the play session, how often participants thought about time, and the passage of time were not significantly different between the two groups. Neither emotional and arousal change nor bodily awareness change before and after playing *Thumper* were significantly different between groups. That is, due to the fact that non-gamers and gamers played on their respective lower and higher performance levels, there were no group differences in related subjective experience of the game.

Correlations With Total Flow, Absorption, and Fluency Scores

The correlation coefficients among the total flow/absorption/fluency and other subjective measures considering all participants are shown in **Table 1**. Since gamers and non-gamers did not differ significantly in subjective variables, the correlation coefficients and related *p* values for the separate groups are presented in the **Supplementary Data** in **Supplementary Table 2**. As shown in **Table 1**, the more flow participants experienced, the less they thought about time ($r = -0.481, p = 0.001$; FDR corrected), and the higher their total game scores was ($r = 0.466, p = 0.002$, FDR corrected). The more absorbed participants were in the game, the less they thought about time ($\rho = -0.530, p < 0.001$, FDR corrected), and the faster time passed for them ($\rho = 0.423, p = 0.005$, FDR corrected).

The last two correlations were driven mainly by the gamers. A positive association was also identified between the absorption score and measures of performance. More absorption in the game led to a higher final gaming score ($\rho = 0.452, p = 0.002$, FDR corrected) and lower total errors ($\rho = -0.377, p = 0.013$). Only two correlations appear to be significant on a 1% alpha level for FSS–Fluency. The more fluency participants felt, the less they thought

TABLE 1 | Correlation coefficients and related *p*-Values between the total flow, absorption, and fluency subscales of flow and related variables for all subjects.

Measures & related variables	Total flow (FSS mean score)	Absorption (FSS-Absorption)	Fluency (FSS-Fluency)
	All <i>r</i> (<i>p</i>)	All <i>r</i> (<i>p</i>)	All <i>r</i> (<i>p</i>)
Thinking about time (STSS)	-0.481^{rho} (0.001^{FDR})	-0.530^{rho} (< 0.001^{FDR})	-0.394^{rho} (0.009)
Speed of time passage (STSS)	0.271 (0.079)	0.423^{rho} (0.005^{FDR})	0.123 (0.433)
Estimated duration of play session (STSS)	0.101 ^{rho} (0.519)	0.069 ^{rho} (0.661)	-0.075 ^{rho} (0.635)
Total final score in the game	0.466 (0.002^{FDR})	0.452^{rho} (0.002^{FDR})	0.390 (0.010)
Total final error in the game	-0.343 (0.024)	-0.377^{rho} (0.013)	-0.232 (0.135)
Bodily awareness (STSS)	-0.160 ^{rho} (0.307)	-0.167 ^{rho} (0.285)	-0.149 ^{rho} (0.342)
Arousal change (SAM-arousal) ¹	-0.032 ^{rho} (0.839)	0.122 ^{rho} (0.437)	-0.141 ^{rho} (0.368)
Valence change (SAM-valence) ¹	0.112 ^{rho} (0.474)	0.213 ^{rho} (0.170)	0.019 ^{rho} (0.904)

^{rho} Spearman correlation results.

¹ Difference between after and before the game session; after-before; significant correlations on the 5% level are marked in bold; FDR = Significant after false-discovery-rate (FDR) adjustment.

about time ($\rho = -0.394$, $p = 0.009$; not significant after FDR), and the higher their final game score was ($r = 0.390$, $p = 0.010$; not significant after FDR).

The Heart-Evoked Potential Measure

Since HEP modulations have been reported in widely distributed scalp electrodes (frontal, central, and parietal sites), as well as a range of latencies between 200 and 600 ms after the R-peak (Pollatos and Schandry, 2004; Babo-Rebelo et al., 2016; Park et al., 2016; Petzschner et al., 2019), we performed a whole brain, whole time-window analysis of the HEPs for the three conditions (pre-game, game, and post-game) using non-parametric permutation tests with 1000 randomizations. Across all subjects, the HEPs significantly differed in the game condition compared to the pre-game and the post-game conditions (considering $\alpha < 0.05$ with FDR correction) over the fronto-central sensors (Fz, FC1, FC2, Cz) in the time window of 400–500 ms after the R-peak (**Figure 2A**). The grand average fronto-central HEP waveform is shown for the three experimental conditions in **Figure 2B**. A *post hoc* analysis revealed that the average HEP amplitude in the time window of 400–500 ms after the R-peak was significantly lower (considering $\alpha < 0.05$ with FDR correction) over the fronto-central sensors for the game condition compared to the pre-game ($p < 0.001$) and the post-game ($p < 0.001$) conditions (**Figures 2B,C**). The lower HEP amplitude during the gameplay represents a lower neural response to the heartbeats while playing, which corresponds with our first hypothesis. No significant difference was found for the HEP waveform between the pre-game and the post-game conditions.

The same suppression of the fronto-central HEP amplitude during playing the game was observed within each group. The topographical maps of the average HEP amplitude over the 400 to 500 ms time window after the R-peak for each condition, as well as the corresponding statistical *p* values (derived from non-parametric permutation tests with 1000 randomizations), are separately presented for each group in **Supplementary Figure 1a** of the **Supplementary Data**. As shown in **Supplementary Figure 1b**, during the game condition the mean HEP amplitude over the 400–500 ms time window in the gamers was slightly higher compared to the non-gamers,

which was not statistically significant ($t(41) = -1.637$, $p = 0.109$). To test whether the observed difference in the HEP amplitude was truly time-locked to the heartbeats, the same statistical analysis was conducted repeatedly (100 times), but this time using surrogate R-peaks, which had the same inter-beat interval and variability as the original R-peaks, but were shifted randomly in time. A surrogate analysis performed separately for each group showed no such significant difference as originally found for the real R-peaks, confirming the fact that the differential HEP during the gameplay compared to the resting state is truly locked to the heartbeats.

To test possible associations between the HEP amplitude and flow measures, the mean amplitude of the HEP while playing the game was calculated within the time-window in which a significant difference was detected. No significant association between the HEP amplitude and the level of flow (FSS mean score) was identified across all subjects ($r = 0.253$, $p = 0.102$). The higher the HEP amplitude, the stronger the absorption in the game ($\rho = 0.377$, $p = 0.013$). This association (**Figure 2D**) which was not significant within each group (**Supplementary Table 4**) associated higher absorption in the game with higher cortical processing of cardiac afferents.

Neural Sources of the Observed Differential Heart-Evoked Potential for the Game and the Pre-game Conditions

We reconstructed neural sources of the HEP signals for the pregame and the game conditions to identify the cortical regions contributing to the HEP effect observed over the fronto-central scalp electrodes. After identifying the time window for which a significant difference between these two conditions was found in sensor space, a cluster-based permutation test can be applied on the reconstructed cortical currents averaged over that time window. The difference in cortical currents between the pre-game and the game conditions was significant over two large regions, one pronounced over the left supplementary motor area (SMA; BA 6) and left primary motor cortex (BA 4) extending to the left primary somatosensory cortex (BA1, 2,3) and the left posterior cingulate cortex (PCC; cluster size = 1928 vertices, Monte Carlo $p = 0.002$).

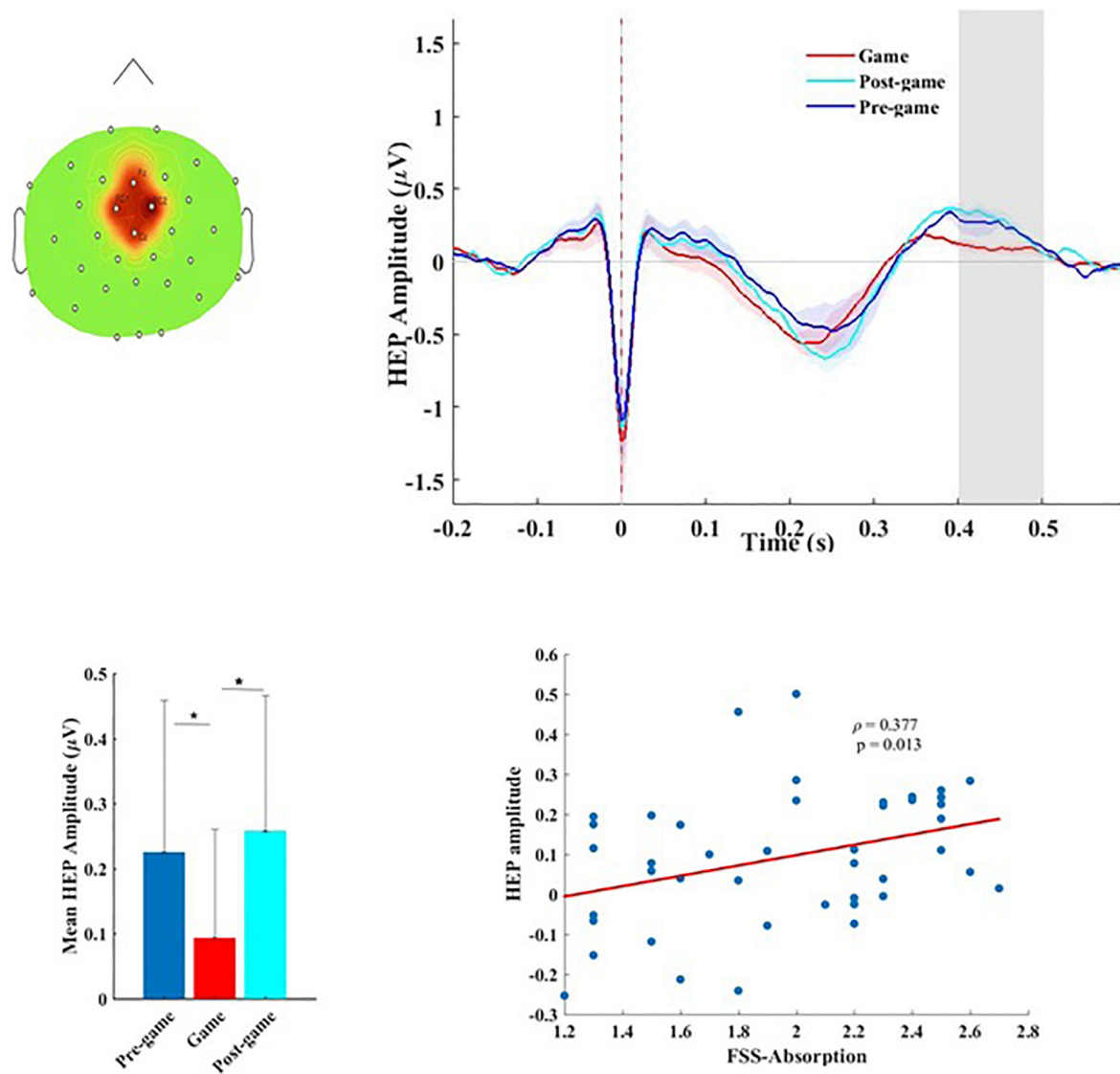
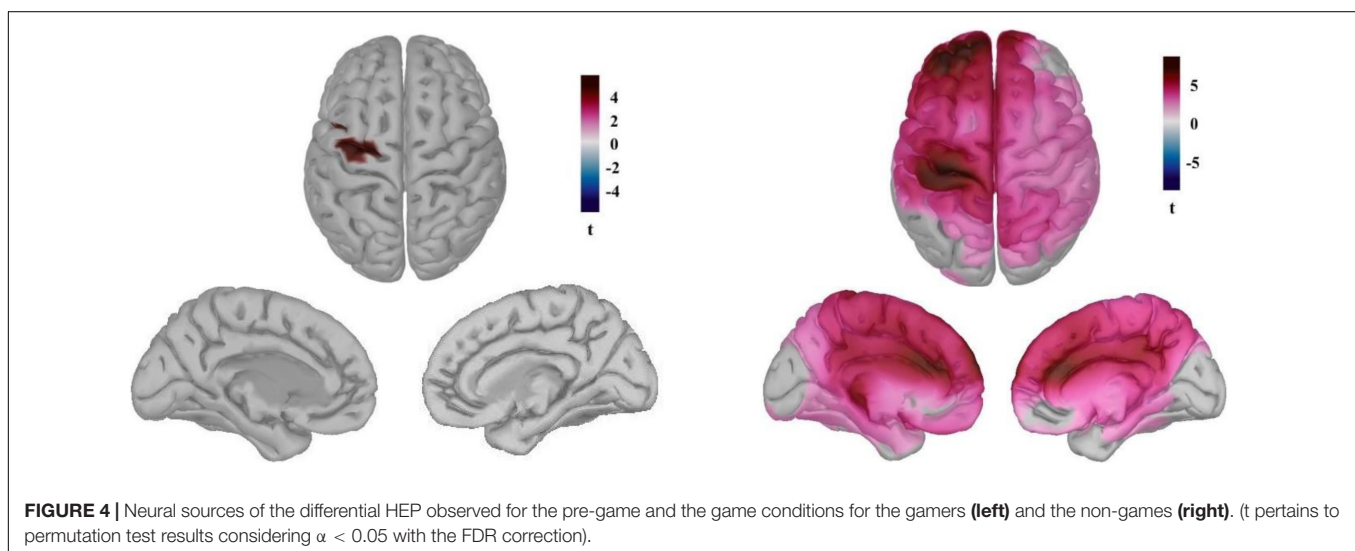
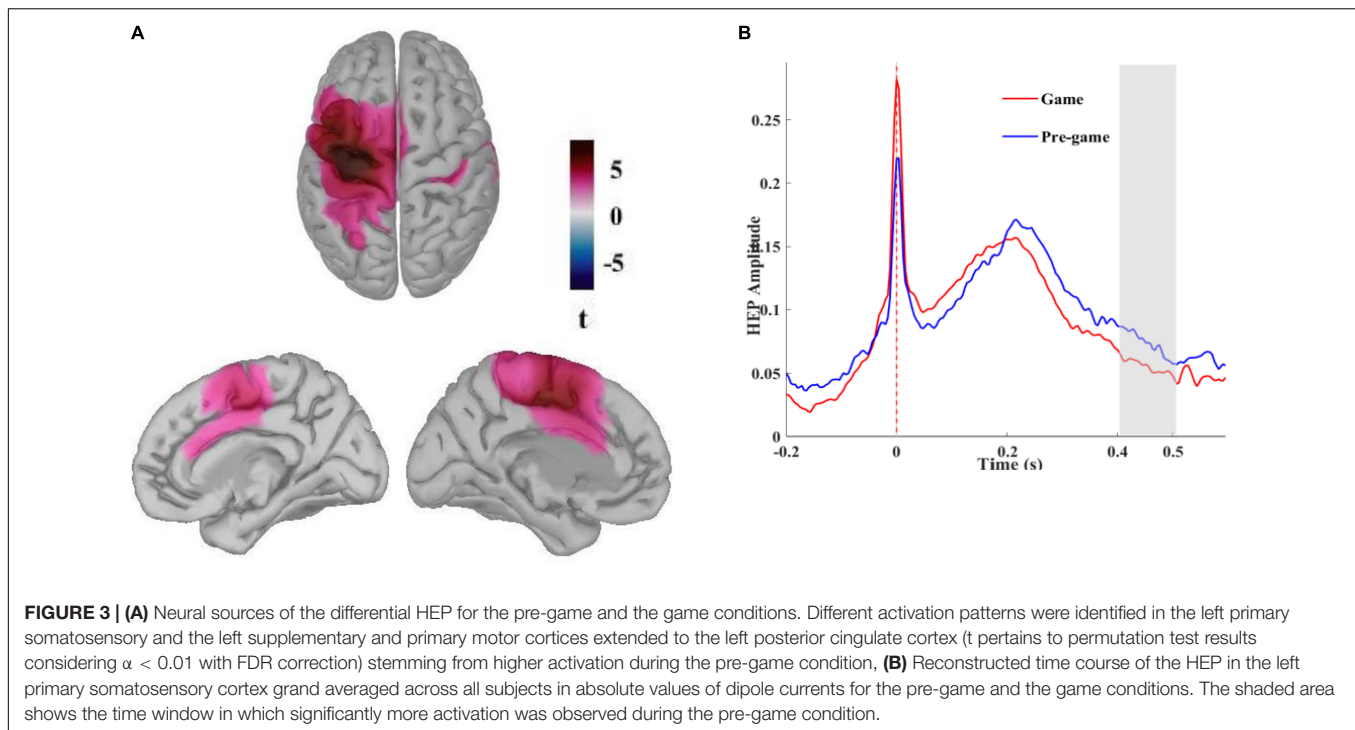


FIGURE 2 | (A) Topographical map of the scalp sensors where the distinct heart-evoked potential (HEP) between the game and the pre-/post-game conditions was observed. The red-colored region including FC1, FC2, Fz, and Cz sensors indicates the location of the electrodes contributing to this significant difference, **(B)** the grand average HEP across fronto-central sensors for the three conditions. The shaded area shows the time window in which we observed a significant difference. **(C)** Mean amplitude and the standard deviation of the HEP over the 400–500 ms time window in which the significant difference was observed for the three conditions, **(D)** Flow-absorption score as a function of the HEP amplitude.

The other cluster included the right primary somatosensory cortex and the right primary motor cortex (BA4) along with the right PPC (cluster size = 1478 vertices, Monte Carlo $p = 0.002$). We also performed a non-parametric permutation test with FDR correction considering $\alpha < 0.01$ to identify the most significant regions. As presented in **Figure 3A**, the HEP neural sources differed significantly between the pre-game and the game condition over the left supplementary and primary motor cortices, the left primary somatosensory cortex extended to the left PCC (**Figure 3A**). Small regions in the right primary motor cortex, the right primary somatosensory cortex, and the right posterior cingulate were also differently

activated between the pre-game and the post-game conditions. The superposition of the statistically different regions with the difference in the absolute values of the cortical currents between the pre-game and the game conditions showed that the activity in the above-mentioned regions during 400 to 500 ms after the R peak was significantly higher in the pre-game condition as compared to the game condition (a video clip demonstrating this source activity difference between the pre-game and the game condition can be found in **Supplementary Materials**).

The time course of the reconstructed neural currents from the primary somatosensory region (**Figure 3B**) demonstrated lower



neural activity locked to the heartbeats in this brain region while playing the game compared to the pre-game condition. We then explored neural sources of the differential HEP observed for the pre-game and the game conditions separately for each group (**Figure 4**). The permutation test results with an FDR correction considering $\alpha < 0.05$ showed that in the gamer group only the supplementary motor cortex (the significant region peaked at MNI coordinates $-25, -14, 60$ with peak $t = 5.81$) was less activated during the game condition compared to the pre-game condition. In non-gamers, though, the left supplementary and primary motor cortex, the left primary somatosensory cortex, the left and right posterior and anterior cingulate cortices, and the

left frontopolar prefrontal cortex (BA10; the significant region peaked at MNI coordinates $-31, 52, 25$ with peak $t = 8.67$) were less activated during the game condition than during the pre-game condition.

Dissimilar neural sources of the observed differential HEP for the gamers and non-gamers may reflect differences in resource allocation to the cortical processing of the heartbeats between these two groups while playing the game vs. the resting state. Similar to the sensor space, the neural sources of the HEP between gamers and non-gamers did not statistically differ in either the pre-game or in the game condition. We also looked at possible correlations between the average neural currents of

the identified sources around 400–500 ms after the R-peak in the game condition and subjective flow measures. No associations were identified across the subjects.

The Function Between the Heart-Evoked Potential Amplitude and Enhanced Automaticity

Difference in resource allocation to the cortical processing of the cardiac information between two groups while playing the game vs. the resting state raised this question whether the observed relation between the HEP amplitude and absorption was related to the enhanced automaticity experienced during flow state. To address this, we extracted the parietal alpha and explored possible correlations between this measure and the HEP amplitude. According to a previous study (Gevins et al., 1997), higher automaticity leads to higher parietal deactivation (higher parietal alpha). The amplitude of the HEP showed a tendency to associate positively with the parietal alpha power ($r = 0.314$, $p = 0.04$); the higher the HEP amplitude, the higher the parietal alpha power. This finding may suggest that the higher absorption in the game resulted in an increased response to the heartbeats (higher HEP amplitude) through enhanced automaticity.

The Function Between the Heart-Evoked Potential Amplitude and Cardiovascular Activity

To explore whether the HEP amplitude is associated with the change in cardiovascular parameters, we searched for possible correlations between the average HEP over the 400–500 ms time window at the fronto-central scalp electrodes and the IBI, the HF- and LF-HRV while playing the game. No associations were identified between the HEP amplitude in this time window and the mentioned cardiac measures during the entire gameplay session. The HEP amplitude was positively associated with the LF- and the HF-HRV power difference between the last and the first 5 min of gameplay ($r = 0.443$, $p = 0.005$; $r = 0.354$, $p = 0.029$, respectively), i.e., the higher the parasympathetic activity recorded during the last 5 min as compared to the first 5 min, the larger the HEP amplitude was.

Peripheral Measures

All physiological measures were extracted and analyzed for the three experimental conditions of the pre-game (5 min), the game (25 min), and the post-game (5 min). The mean value of the neurophysiological measures for the three experimental conditions and for each group along with the statistical results are presented in the **Supplementary Table 3**. To examine the pattern of cardiorespiratory activity more precisely during the whole 25-min gameplay, we segmented the recorded physiological signals of the game condition into non-overlapping, 5-min intervals and analyzed the differences between corresponding cardiovascular and respiratory measures for the first and the last intervals. Here we reported the findings regarding all participants. For some differences between gamers and non-gamers we refer to the **Supplementary Data (Supplementary Table 4)**.

Respiratory Activity

We analyzed two parameters from the respiration signals: the mean respiration rate (RR-mean), the standard deviation of the

respiration rate (RR-STD). The Greenhouse-Geisser correction revealed a significant main effect of condition (pre-game, game, post-game) for the RR-mean ($F(1.25, 51.28) = 36.70$; $p < 0.001$, $\eta^2 = 0.203$) and RR-STD ($F(1.55, 63.64) = 42.602$; $p < 0.001$, $\eta^2 = 0.314$). *Post hoc* tests showed that the mean and variance of the respiration rate were significantly higher during the game compared to the pre-game ($p < 0.001$) and the post-game ($p < 0.001$) conditions. The main group effect (gamers, non-gamers), as well as the interaction between condition and group, was not significant for any of these parameters (see **Supplementary Table 4**).

There was no association between the RR-mean and flow measures during the whole gameplay session. The respiration-rate difference between the last 5 min and the first 5 min of gameplay was negatively correlated with absorption across all subjects ($\rho = -0.424$, $p = 0.005$). A slower respiration rate at the end of the gameplay session compared to the beginning (lower sympathetic and higher parasympathetic activity) is related to greater absorption in the game. The variation in the respiration rate (RR-STD) during the whole 25-min game interval was significantly correlated with the total flow ($r = 0.418$, $p = 0.005$), and the absorption sub-scale of flow ($\rho = 0.505$, $p < 0.001$). These two correlations which were driven mainly by the gamers (**Supplementary Table 4**) highlighted that the more variation in the respiration rate was associated with more flow and a stronger absorption in the game. The difference in the RR-STD value between the last and the first 5 min of gameplay showed no significant correlations with flow measures.

Cardiac Activity

The inter-beat interval (IBI) of the ECG signals differed significantly between conditions ($F(2, 82) = 8.295$; $p < 0.001$, $\eta^2 = 0.01$). *Post hoc* testing revealed a smaller IBI (higher heart rate) during gameplay compared to the pre-game ($p = 0.005$) and the post-game ($p < 0.001$) conditions across all subjects. As shown in **Supplementary Figure 2a** in the **Supplementary Data**, the IBI was overall significantly longer for the non-gamers than for the gamers ($F(1, 41) = 4.68$; $p = 0.03$, $\eta^2 = 0.096$) reflecting lower sympathetic activity in non-gamers. The interaction of condition and group was not significant. The data of five participants whose breathing rates were beyond 9–24 cycles per minute (below 9 and above 24 cycles) were excluded from the frequency-domain analysis of the HRV. We found a significant main-condition effect for both the LF-HRV ($F(2, 72) = 6.21$; $p = 0.003$, $\eta^2 = 0.032$) and the HF-HRV ($F(2, 72) = 16.31$; $p < 0.001$, $\eta^2 = 0.060$) measures. *Post hoc* testing showed that both measures were lower during the game compared to the pre-game ($p = 0.047$, $p < 0.001$, respectively) and the post-game ($p = 0.003$, $p < 0.001$, respectively) conditions. No main effect of group or interaction effect was found for these measures.

Figure 5A illustrates that the average IBI continuously declined while playing the game from the first interval up to the third 5-min interval, followed by a plateau for the following 5-min interval, and with a slight increase during the final 5-min interval. During the entire 25 min of gameplay, the shorter the IBI (i.e., the faster heart rate), the higher the total flow score ($r = -0.325$, $p = 0.033$). This effect, which was mainly driven

by the non-gamers (**Supplementary Table 4**), showed that flow was associated with higher sympathetic activity during the whole gameplay period.

Correlations between IBI and absorption/fluency were not significant. The IBI difference between the last and the first 5 min of gameplay was significantly correlated with flow ($r = 0.435$, $p = 0.003$), absorption ($\rho = 0.432$, $p = 0.004$), and fluency ($r = 0.354$, $p = 0.020$). This relationship, which is mostly driven by non-gamers (**Supplementary Table 4**), emphasizes that a higher IBI and consequently slower heart rate (lower sympathetic activity) during the last 5 min of playing compared to the first 5 min is positively associated with higher levels of flow (**Figure 5B**), absorption (**Figure 5C**), and fluency.

The LF and HF-HRV measures were significantly decreased during gameplay compared to the pre-game and the post-game resting states (see **Supplementary Figure 2b**). During gameplay, there were no significant associations between LF-HRV and HF-HRV during the game and the total flow experience, absorption, fluency. The HF-HRV power difference between the last and the first 5-min interval of gameplay (**Supplementary Figure 2a**) showed a positive correlation with the FSS-absorption score ($\rho = 0.365$, $p = 0.024$). Higher HF-HRV power (higher parasympathetic activity) during the last 5 min compared to the first 5 min of gameplay is associated with higher absorption in the game. No significant correlations were found for the HF-HRV difference and flow or fluency scores. A similar trend for an association was found for the LF-HRV power values showing higher LF-HRV power during the last 5-min interval of gameplay compared to the first, which was associated with higher absorption in the game ($\rho = 0.353$, $p = 0.029$). These findings show that, although playing the game results in a reduction in parasympathetic activity (lower LF- and HF-HRV), the increase in parasympathetic activity (or, more precisely, less inhibition of parasympathetic activity) at the end of the game leads to higher levels of absorption.

DISCUSSION

In the present study, we investigated the experience of flow, absorption, and subjective time in gamers and non-gamers while playing the video game *Thumper*. We concentrated on two important aspects of the flow experience, namely the loss of self-referential processing and increased arousal. We used the HEP amplitude as an index of cortical processing of cardiac afferents to evaluate self-referential processing during flow. Associations between the activity of the autonomous nervous system and flow were assessed by evaluating cardiorespiratory measures, including the mean and standard deviation of the respiration rate, the mean IBI, and the average LF- and HF-HRV power.

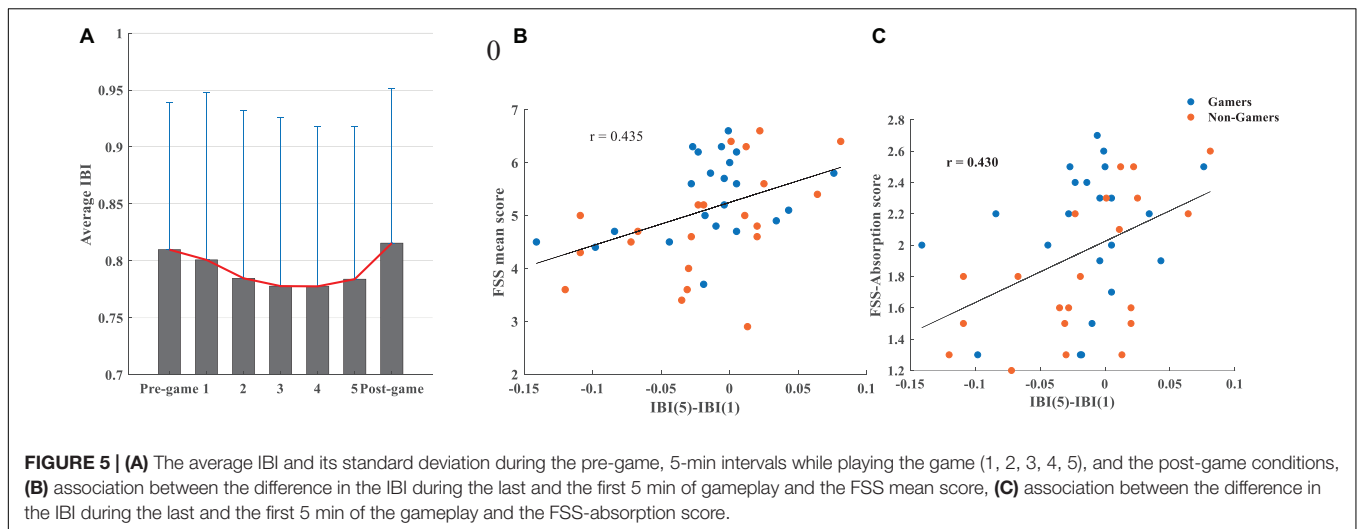
Our behavioral findings showed that playing *Thumper* elicited comparable levels of flow in both groups with no significant differences in terms of total flow, absorption, or fluency scores. Also, the measures of subjective time did not differ between the groups. We interpret these findings in the way that playing the video game *Thumper* on a level chosen by the participant according to their skill facilitated flow induction. In this way,

non-gamers tended to play at relatively lower levels whereas gamers chose relatively higher levels, which made it possible for each group to achieve flow.

The experience of flow was related to the subjective experience of time. The higher the participants' flow score and the higher the degrees of absorption and fluency they experienced during the game, the less they thought about time. Higher absorption scores were also positively associated with a faster passage of time. These results coincide with those of previous studies that found an association between the experience of flow and subjective time (Keller and Bless, 2008; Rutrecht et al., 2021). Experienced flow, absorption, and fluency were positively correlated with performance, as reflected in the attained final score. This positive association was also reported by several other studies assessing flow (Engeser and Rheinberg, 2008; Jin, 2012; Yun et al., 2017; Rutrecht et al., 2021) and aligns well with the notion of peak performance resulting from the flow experience (Landhäuser and Keller, 2012; Khoshnoud et al., 2020). Task engagement has been related to performance (Matthews et al., 2010). Flow as a positively toned state of high task engagement, signals a cognitive-adaptive process (Matthews et al., 2010) that leads to more effective mobilization of attentional resources to the task (De Sampaio Barros et al., 2018).

Flow Is Linked to a More Active Brain–Heart Interaction

The HEP amplitude significantly decreased 400–500 ms after the R-peak over the fronto-central scalp electrodes while participants played the game as compared to the pre-game and the post-game conditions. The same HEP difference was observed separately for gamers and non-gamers in the same time range. The use of surrogate R-peaks confirmed that the observed differential HEP during the gameplay session compared to the resting states was actually locked to the heartbeats. This HEP amplitude difference can be considered free from artifactual components because confounding noises, such as cardiac field artifacts, were identified and removed in the preprocessing step using the AMICA. The differential HEP was observed during the time period (400 ms after the R-peak) which is known to be less affected by the cardiac field artifact (Park and Blanke, 2019). The location and latency of the observed differential HEP coincide with previous literature reporting differential HEP between 250 and 500 ms after the R-peak over the fronto-central scalp sensors in modulated self-related processing (Park et al., 2016; Sel et al., 2017; Babo-Rebelo et al., 2019, 2016), and in mental disorders associated with atypical states of self-consciousness, such as depression (Schulz et al., 2015). This lower cortical response to the cardiac afferent while playing may reflect lower self-awareness during gameplay compared to the resting state, in which the sense of self is comparably higher, a typical experience in waiting situations (Jokic et al., 2018). The fact that the fronto-central HEP amplitude was significantly higher during the resting states before and after the game as compared to the game condition reflects higher self-referential processing while waiting. As we have empirically shown, a real waiting period of several minutes is related to stronger negative self-awareness (i.e., boredom) and



a slower felt passage of time (Witowska et al., 2020; Alvarez Igarzábal et al., 2021). A waiting period without distraction can be defined as a state with higher levels of self-awareness. The higher HEP amplitude during resting states suggests that bodily signals (such as from the heart) are more readily available during waiting time. This finding corresponds also with a previous study by Wei et al. (2016) that reported a lower fronto-central HEP amplitude during an exteroceptive state (eyes-open resting state) compared to an interoceptive state (eyes-closed resting state). In this regard, playing the game as an intense exteroceptive state led to the lower HEP amplitude compared to the resting states (pre- and post-game), which can be considered relatively more interoceptive.

During the gameplay session, we found a positive correlation between the HEP amplitude and the level of absorption. Individuals who were more absorbed while playing the game exhibited higher HEP amplitudes. Higher levels on the absorption subscale of the flow questionnaire by definition are a strong indicator of flow (Csikszentmihalyi, 1975; Peifer et al., 2014). As formulated by Csikszentmihalyi (1975, p. 39), during flow “your concentration is very complete. Your mind isn’t wandering, you are not thinking of something else; you are totally involved in what you are doing.” Therefore the positive correlation between the absorption subscale of FSS and the HEP amplitude shows a positive association between flow and the HEP. This could be interpreted in the way that flow-absorption during the game necessitated more information from the body, leading to a stronger connection between the brain and the heart. HEP amplitudes around 400–500ms after the ECG R-peak have been mostly investigated and interpreted as correlates of the successive cognitive elaboration of interoceptive information (Baranauskas et al., 2017). In the context of interoceptive predictive coding concepts (Seth et al., 2012; Critchley and Garfinkel, 2018), the brain has been shown to use the interoceptive signals (cardiac information) to successfully predict upcoming exteroceptive events (Park et al., 2014; Pfeiffer and De Lucia, 2017; Banellis and Cruse, 2020). *Thumper* is a dynamic rhythm game that demands high levels of concentration

while performing challenging sensorimotor tasks. It is reasonable therefore to assume that precise prediction of sensory stimuli and adequate reactions to them are dependent on information from body states and vice versa, especially since navigating in *Thumper* necessitates the continuous anticipation of upcoming objects the player has to synchronize with. Therefore, we suggest that the stronger HEP response in more absorbed individuals stems from a more efficient brain-body interaction. This interpretation is substantiated by our findings that more absorbed participants had a higher total final score and a lower total final error in the game (Table 1).

Alpha oscillations has been reported to be important for the efficiency of brain–heart interaction (Luft and Bhattacharya, 2015). Our result showed that higher HEP amplitude is associated with higher parietal alpha power (more deactivation of the parietal cortex). Higher automaticity achieved with increased practice on the task has been linked to higher parietal alpha oscillations (Gevins et al., 1997). We argue that enhanced automaticity during flow may also contribute to the observed function between the HEP amplitude and absorption. The enhanced automaticity during moments of flow through increased reliance on implicit information-processing resources (Dietrich, 2004) may facilitate brain–heart interaction during flow by making more frontal and parietal resources available for the processing of cardiac afferents. It has been reported that enhanced automaticity during flow helps individuals maintain a sustained level of attention so that the increased task demands and challenges can be carried out without a further increase in felt attentional effort (Harris et al., 2017; Yelamanchili, 2018; Khoshnoud et al., 2020).

The Heart-Evoked Potential Effect Generated in Regions Associated With Interoceptive Processing

We demonstrated that the HEP effect was generated in regions associated with the processing of internal bodily signals across all participants. The identified neural sources contributing to the observed differential fronto-central HEP between the game and the pre-game conditions covered the left SMA and the left

primary motor cortex, extending to the primary somatosensory cortex and the PCC. The activity in these regions was significantly lower during gameplay compared to the pre-game resting state. Since the primary somatosensory cortex and the PCC are two important regions known for the processing of interoceptive signals (Kern et al., 2013; Park and Blanke, 2019), the lower HEP amplitude observed during gameplay corresponds with the lower cortical source currents in regions known for the processing of cardiac signals.

The HEP effect was localized differently in the brain for gamers and non-gamers. While for gamers only the SMA was less activated during the game compared to the pre-game condition, for non-gamers an extended area including the left SMA, the left primary motor cortex, the left primary somatosensory cortex, the left and right PCC and anterior cingulate cortex (ACC), and the left frontopolar prefrontal cortex were less activated during gameplay compared to the pre-game condition. An explanation for this huge difference between gamers and non-gamers in size and location of the identified neural sources contributing to the observed HEP effect may be related to the enhanced skill and automaticity of gamers during the gameplay. For gamers, playing *Thumper* was not such a demanding challenge due to their experience, therefore the HEP source difference before and during the game was relatively smaller. For the non-gamers *Thumper* posed a higher challenge, as they lacked the experience and skills of gamers. Consequently, a larger difference in neural sources was detected between the resting state before the game and the game condition in non-gamers. Dietrich (2004) proposed that higher automaticity during moments of flow can be achieved by increased reliance on implicit information processing (which is supported primarily by the basal ganglia) instead of explicit information processing (which is supported by the frontal and parietal lobes) in the brain. Considering this, the higher automaticity in gamers achieved through the use of more implicit information processing may have allowed them to allocate comparable resources to the cortical processing of the heartbeats during gameplay to the resting state before the game. Non-gamers exhibited less automaticity during the game because of their lack of skills and relied more on the explicit information-processing system (which is based on frontal and parietal resources). This may result in fewer available frontal and parietal resources for the processing of cardiac information while playing the game. Thus, the extended brain regions in frontal and parietal cortex were less activated during gameplay compared to the pre-game condition in these participants in response to the heartbeats.

Contribution of Parasympathetic Activity to the Observed Heart-Evoked Potential Effect

Several studies have shown a link between emotional and arousal processes and the HEP amplitude (Fukushima et al., 2011; Shao et al., 2011; Schulz et al., 2013; Luft and Bhattacharya, 2015), but have not always observed a difference in cardiac parameters between experimental conditions in which HEP differences were found (Fukushima et al., 2011; Park and Blanke, 2019). Considering the whole gameplay session, we did not observe

any associations between the cardiorespiratory measures and the HEP amplitude. The HEP amplitude was positively correlated with the difference in parasympathetic activity (higher LF and HF-HRV) between the last 5 and the first 5 min of gameplay. This observed positive association shows that the brain-heart interaction is associated with the activation of the vagus nerve at the end of the game compared to the beginning. It has been stated that the vagus nerve is activated during self-transcendent positive emotions (Kitson et al., 2020).

Flow Is Associated With Increased Sympathetic and Less Inhibition of Parasympathetic Activity

The total flow and absorption levels were positively correlated with variations in the respiration rates (RR-STD), revealing greater variations during the game, higher flow, and stronger absorption in the game. Respiration has been shown to contribute to the short-term modulation of the sympathetic nervous system (Narkiewicz et al., 2006) with a faster respiration rate associated with higher sympathetic activity (Wientjes, 1992; de Manzano et al., 2010). A more varied respiration rate while playing the video game shows greater short-term modulations of the sympathetic nervous system, which may help the respiration system to co-vary with the demands of the game and thereby facilitate the experience of flow and absorption. Total flow was also positively correlated with the average IBI values during the entire game session. The higher the total flow participants reported, the faster their heart rate. These findings align well with the reported positive association between increased sympathetic activity and flow in previous studies (Kivikangas, 2006; de Manzano et al., 2010; Bian et al., 2016). The lower mean-RR (slower respiration rate) and the longer IBI (slower heart rate) during the last 5 min compared to the first were associated with higher absorption in the game. The IBI difference between the last and the first 5-min intervals was also positively correlated with the total flow and fluency levels. These results indicate that higher parasympathetic activity during the last 5 min of gameplay compared to the first 5 min led to the higher flow and absorption levels.

In contrast to studies reporting a relationship between flow and HRV measures (Chanel et al., 2011; Keller et al., 2011; Harmat et al., 2015; Harris et al., 2016; De Sampaio Barros et al., 2018; Kozhevnikov et al., 2018), no significant associations were identified between LF-HRV/HF-HRV and flow measures during the 25 min of gameplay. Parasympathetic activity (including LF and HF-HRV) decreased during gameplay with no significant difference between gamers and non-gamers. This general inhibition of parasympathetic activity reflects increased mental effort while playing as compared to the pre- and post-game resting states. The higher LF-HRV and HF-HRV during the last 5 min of gameplay compared to the first 5 min were positively correlated with the level of absorption in the game among all participants. The reduced inhibition of parasympathetic activity at the end of the game led to higher absorption.

Combining these findings and considering the whole gameplay session, we suggest that flow is associated with increased sympathetic activity manifested by more variable respiration and a faster heart rate. Participants who had less inhibition of parasympathetic activity at the end of the gameplay session compared to the beginning were better able to actually control the heightened sympathetic activity and consequently reported higher flow, absorption, and fluency. According to Porges' (2001) polyvagal theory, parasympathetic influences are essential for an individual's successful adaptation to changing environmental demands. Studies have related parasympathetic activity (specifically HRV) to working- memory performance, mental workload, and attention (Veltman and Gaillard, 1998; Hansen et al., 2003). Hansen et al. (2003) reported that a higher resting state parasympathetic activity (higher HF-HRV) is associated with better performance in a working-memory task and continuous performance test. Both increased sympathetic activity and decreased parasympathetic activity have been associated with lower performance (Forte et al., 2019) in a systematic review. Considering the identified positive association between flow and sympathetic activity in this study, relatively less inhibition of parasympathetic activity at the end of the gameplay session compared to the beginning may indicate a higher ability to respond flexibly to the changing demands of the game. Our findings align well with the reported link between the co-activation of sympathetic and parasympathetic activity and the flow experience (de Manzano et al., 2010; Harmat et al., 2015; Tian et al., 2017). The co-activation in our results was demonstrated by increased overall sympathetic activity and decreased inhibition of parasympathetic activity during the last 5 min of gameplay.

All in all, our findings indicate that game-induced flow modulates subjective time perception in terms of less thinking about time and the feeling of a faster passage of time. We showed that the monitoring of cardiac afferent information being processed by the brain can serve as an objective measure to assess the level of absorption in the game. The positive association identified between the fronto-central HEP amplitude and absorption during the game can be interpreted by a better brain–heart interaction that led to a better performance, i.e., a higher final score and a lower total error. Our findings also provide evidence for the relationship between the co-activation of the sympathetic and parasympathetic nervous systems and the flow experience. As an applied perspective of our work, the HEP amplitude could function as a neural marker of flow-absorption in patients with depression and anxiety who have lost their ability to lose track of the self and time. It is known that individuals with Major Depressive Disorder have deficits in emotional self-regulation and also report a drastically slowed down subjective passage of time (Vogel et al., 2018). Further inquiries are needed to assess whether states of flow elicited by games like Thumper ameliorate symptoms of depression by accelerating the subjective flow of time. In one earlier study by Kühn et al. (2018), depressed individuals played the video game Boson X, which

has similar characteristics to *Thumper*, for 6 weeks and this reduced rumination and enhanced cognitive abilities. The HEP amplitude thereby could function as neural marker for increased states of flow-absorption during tasks individuals with depression are undertaking. It could reflect improvements of their psychological state.

DATA AVAILABILITY STATEMENT

The dataset generated for this study can be found on the Open Science Framework website, https://osf.io/djre6/?view_only=1bbd140e80b24c45ac2f8c25123c6d51.

ETHICS STATEMENT

This study was reviewed and approved by the Local Ethics Committee of the Institute for Frontier Areas of Psychology and Mental Health (IGPP_2019_01).

AUTHOR CONTRIBUTIONS

SK performed data curation, conducted the analysis, and wrote the first draft. All authors contributed to conceptualization, design of the study, manuscript revision, and approved the submitted version.

FUNDING

This study was funded by the EU, Horizon 2020 Framework Program, FET Proactive (VIRTUALTIMES consortium, grant agreement Id: 824128 to MW). VIRTUALTIMES — Exploring and modifying the sense of time in virtual environments — includes the following groups with the principal investigators Kai Vogeley (Cologne), MW (Freiburg), Anne Giersch (Strasbourg), Marc Erich Latoschik, Jean-Luc Lugin (Würzburg), Giulio Jacucci, Niklas Ravaja (Helsinki), Xavier Palomer, and Xavier Oromi (Barcelona).

ACKNOWLEDGMENTS

The authors would like to thank David Leitritz for his assistance in data recording.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2022.819834/full#supplementary-material>

REFERENCES

- Alvarez Igarzábal, F. (2019). *Time and Space in Video Games: A Cognitive-Formalist Approach*. Bielefeld: Transcript verlag.
- Alvarez Igarzábal, F., Hruby, H., Witowska, J., Khoshnoud, S., and Wittmann, M. (2021). What happens while waiting in virtual reality? A comparison between a virtual and a real waiting situation concerning boredom, self-regulation, and the experience of time. *Technol. Mind Behav.* 2, 1–11. doi: 10.1037/tmb0000038
- Azzalini, D., Rebollo, I., and Tallon-Baudry, C. (2019). Visceral Signals Shape Brain Dynamics and Cognition. *Trends Cogn. Sci.* 23, 488–509. doi: 10.1016/j.tics.2019.03.007
- Babo-Rebelo, M., Buot, A., and Tallon-Baudry, C. (2019). Neural responses to heartbeats distinguish self from other during imagination. *NeuroImage* 191, 10–20. doi: 10.1016/j.neuroimage.2019.02.012
- Babo-Rebelo, M., Richter, C. G., and Tallon-Baudry, C. (2016). Neural responses to heartbeats in the default network encode the self in spontaneous thoughts. *J. Neurosci.* 36, 7829–7840. doi: 10.1523/JNEUROSCI.0262-16.2016
- Banellis, L., and Cruse, D. (2020). Skipping a Beat: heartbeat-Evoked Potentials Reflect Predictions during Interoceptive-Exteroceptive Integration. *Cereb. Cortex Commun.* 1, 1–14. doi: 10.1093/texcom/tgaa060
- Baranauskas, M., Grabauskaitė, A., and Griškova-Bulanova, I. (2017). Brain responses and self-reported indices of interoception: heartbeat evoked potentials are inversely associated with worrying about body sensations. *Physiol. Behav.* 180, 1–7. doi: 10.1016/j.physbeh.2017.07.032
- 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
- Bian, Y., Yang, C., Gao, F., Li, H., Zhou, S., Li, H., et al. (2016). A framework for physiological indicators of flow in VR games: construction and preliminary evaluation. *Pers. Ubiquitous Comput.* 20, 821–832. doi: 10.1007/s00779-016-0953-5
- Bisson, N., Tobin, S., and Grondin, S. (2012). Prospective and retrospective time estimates of children: a comparison based on ecological tasks. *PLoS One* 7:e33049. doi: 10.1371/journal.pone.0033049
- Bradley, M. M., and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* 25, 49–59. doi: 10.1016/0005-7916(94)90063-9
- Cacioppo, J. T., Tassinary, L. G., and Berntson, G. (2007). *Handbook of psychophysiology*. Cambridge: Cambridge University press.
- Chanel, G., Rebetez, C., Bétrancourt, M., and Pun, T. (2011). Emotion Assessment From Physiological Signals for Adaptation of Game Difficulty. *IEEE Trans. Syst. Man Cybern. A Syst. Hum.* 41, 1052–1063. doi: 10.1109/TSMCA.2011.2116000
- Chang, C. Y., Hsu, S. H., Pion-Tonachini, L., and Jung, T. P. (2020). Evaluation of Artifact Subspace Reconstruction for Automatic Artifact Components Removal in Multi-Channel EEG Recordings. *IEEE Trans. Biomed. Eng.* 67, 1114–1121. doi: 10.1109/TBME.2019.2930186
- Craig, A. D. (2009). Emotional moments across time: a possible neural basis for time perception in the anterior insula. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 1933–1942. doi: 10.1098/rstb.2009.0008
- Critchley, H. D., and Garfinkel, S. N. (2018). The influence of physiological signals on cognition. *Curr. Opin. Behav. Sci.* 19, 13–18. doi: 10.1016/j.cobeha.2017.08.014
- Critchley, H., and Nagai, Y. (2013). “Electrodermal Activity (EDA),” in *Encyclopedia of Behavioral Medicine*, eds M. D. Gellman and J. R. Turner (New York: Springer), 666–669. doi: 10.1007/978-3-642-28753-4_100709
- Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety*. Hoboken, NJ: Jossey-Bass Publishers.
- Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York, NY: Harper Perennial.
- Csikszentmihalyi, M., and Csikszentmihalyi, I. S. (1992). *Optimal experience: Psychological studies of flow in consciousness*. Cambridge: Cambridge University Press. doi: 10.7551/mitpress/9780262013840.003.0010
- Damasio, A., and Carvalho, G. B. (2013). The nature of feelings: evolutionary and neurobiological origins. *Nat. Rev. Neurosci.* 14, 143–152. doi: 10.1038/nrn3403
- de Manzano, Ö., Theorell, T., Harmat, L., and Ullén, F. (2010). The psychophysiology of flow during piano playing. *Emotion* 10, 301–311. doi: 10.1037/a0018432
- De Sampaio Barros, F. M., Araújo-Moreira, F. M., Trevelin, L. C., and Radel, R. (2018). Flow experience and the mobilization of attentional resources. *Cogn. Affect. Behav. Neurosci.* 18, 810–823. doi: 10.3758/s13415-018-0606-4
- Delorme, A., Palmer, J., Onton, J., Oostenveld, R., and Makeig, S. (2012). Independent EEG Sources Are Dipolar. *PLoS One* 7:e30135. doi: 10.1371/journal.pone.0030135
- Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* 31, 968–980. doi: 10.1016/j.neuroimage.2006.01.021
- Dietrich, A. (2004). Neurocognitive mechanisms underlying the experience of flow. *Conscious. Cogn.* 13, 746–761. doi: 10.1016/j.concog.2004.07.002
- Engesser, S., and Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motiv. Emot.* 32, 158–172. doi: 10.1007/s11031-008-9102-4
- Fong, C. J., Zaleski, D. J., and Leach, J. K. (2015). The challenge-skill balance and antecedents of flow: a meta-analytic investigation. *J. Posit. Psychol.* 10, 425–446. doi: 10.1080/17439760.2014.967799
- Forte, G., Favieri, F., and Casagrande, M. (2019). Heart Rate Variability and Cognitive Function: a Systematic Review. *Front. Neurosci.* 13:710. doi: 10.3389/fnins.2019.00710
- Frewen, J., Finucane, C., Savva, G. M., Boyle, G., Coen, R. F., and Kenny, R. A. (2013). Cognitive function is associated with impaired heart rate variability in ageing adults: the Irish longitudinal study on ageing wave one results. *Clin. Auton. Res.* 23, 313–323. doi: 10.1007/s10286-013-0214-x
- Fukushima, H., Terasawa, Y., and Umeda, S. (2011). Association between interoception and empathy: evidence from heartbeat-evoked brain potential. *Int. J. Psychophysiol.* 79, 259–265. doi: 10.1016/j.ijpsycho.2010.10.015
- Gevens, A., Smith, M. E., McEvoy, L., and Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. *Cereb. Cortex* 7, 374–385. doi: 10.1093/cercor/7.4.374
- Goldberg, I. I., Harel, M., and Malach, R. (2006). When the Brain Loses Its Self: prefrontal Inactivation during Sensorimotor Processing. *Neuron* 50, 329–339. doi: 10.1016/j.neuron.2006.03.015
- Gramfort, A., Papadopoulos, T., Olivi, E., and Clerc, M. (2010). OpenMEEG: opensource software for quasistatic bioelectromagnetics. *BioMed. Eng. Online* 9:45. doi: 10.1186/1475-925X-9-45
- Hansen, A. L., Johnsen, B. H., and Thayer, J. F. (2003). Vagal influence on working memory and attention. *Int. J. Psychophysiol.* 48, 263–274. doi: 10.1016/S0167-8760(03)00073-4
- Harmat, L., de Manzano, Ö., Theorell, T., Högman, L., Fischer, H., and Ullén, F. (2015). Physiological correlates of the flow experience during computer game playing. *Int. J. Psychophysiol.* 97, 1–7. doi: 10.1016/j.ijpsycho.2015.05.001
- Harris, D. J., Vine, S. J., and Wilson, M. R. (2016). Is flow really effortless? The complex role of effortful attention. *Sport Exerc. Perform. Psychol.* 6, 103–114. doi: 10.1037/spy0000083
- Harris, D. J., Vine, S. J., and Wilson, M. R. (2017). Neurocognitive mechanisms of the flow state. *Prog. Brain Res.* 234, 221–243. doi: 10.1016/bs.pbr.2017.06.012
- Huskey, R., Craighead, B., Miller, M. B., and Weber, R. (2018). Does intrinsic reward motivate cognitive control? a naturalistic-fMRI study based on the synchronization theory of flow. *Cogn. Affect. Behav. Neurosci.* 18, 902–924. doi: 10.3758/s13415-018-0612-6
- Jin, S. A. A. (2012). “Toward Integrative Models of Flow”: effects of Performance, Skill, Challenge, Playfulness, and Presence on Flow in Video Games. *J. Broadcast. Electron. Media* 56, 169–186. doi: 10.1080/08838151.2012.678516
- Jokic, T., Zakay, D., and Wittmann, M. (2018). Individual Differences in Self-Rated Impulsivity Modulate the Estimation of Time in a Real Waiting Situation. *Timing Time Percept.* 6, 71–89. doi: 10.1163/22134468-00002101
- Ju, U., and Wallraven, C. (2019). Manipulating and decoding subjective gaming experience during active gameplay: a multivariate, whole-brain analysis. *NeuroImage* 188, 1–13. doi: 10.1016/j.neuroimage.2018.11.061
- Keller, J., and Bless, H. (2008). Flow and regulatory compatibility: an experimental approach to the flow model of intrinsic motivation. *Pers. Soc. Psychol. Bull.* 34, 196–209. doi: 10.1177/0146167207310026
- Keller, J., and Blomann, F. (2008). Locus of control and the flow experience: an experimental analysis. *Eur. J. Pers.* 22, 589–607. doi: 10.1002/per.692

- Keller, J., Bless, H., Blomann, F., and Kleinbühl, D. (2011). Physiological aspects of flow experiences: skills-demand-compatibility effects on heart rate variability and salivary cortisol. *J. Exp. Soc. Psychol.* 47, 849–852. doi: 10.1016/j.jesp.2011.02.004
- Kern, M., Aertsen, A., Schulze-Bonhage, A., and Ball, T. (2013). Heart cycle-related effects on event-related potentials, spectral power changes, and connectivity patterns in the human ECoG. *NeuroImage* 81, 178–190. doi: 10.1016/j.neuroimage.2013.05.042
- Khoshnoud, S., Alvarez Igarzábal, F., and Wittmann, M. (2020). Peripheral-physiological and neural correlates of the flow experience while playing video games: a comprehensive review. *PeerJ* 8:e10520. doi: 10.7717/peerj.10520
- Khoshnoud, S., Shamsi, M., Nazari, M. A., and Makeig, S. (2018). Different cortical source activation patterns in children with attention deficit hyperactivity disorder during a time reproduction task. *J. Clin. Exp. Neuropsychol.* 40, 633–649. doi: 10.1080/13803395.2017.1406897
- Kitson, A., Chirico, A., Gaggioli, A., and Riecke, B. E. (2020). A Review on Research and Evaluation Methods for Investigating Self-Transcendence. *Front. Psychol.* 11:547687. doi: 10.3389/fpsyg.2020.547687
- Kivikangas, J. (2006). *Sychophysiology Of Flow Experience: An Explorative Study*. Ph.D. thesis. Finland: University of Helsinki.
- Kozhevnikov, M., Li, Y., Wong, S., Obana, T., and Amihai, I. (2018). Do enhanced states exist? Boosting cognitive capacities through an action video-game. *Cognition* 173, 93–105. doi: 10.1016/j.cognition.2018.01.006
- Kühn, S., Berna, F., Lüdtkke, T., Gallinat, J., and Moritz, S. (2018). Fighting depression: action video game play may reduce rumination and increase subjective and objective cognition in depressed patients. *Front. Psychol.* 9:129. doi: 10.3389/fpsyg.2018.00129
- Laborde, S., Mosley, E., and Thayer, J. F. (2017). Heart rate variability and cardiac vagal tone in psychophysiological research - Recommendations for experiment planning, data analysis, and data reporting. *Front. Psychol.* 8:213. doi: 10.3389/fpsyg.2017.00213
- Landhäuser, A., and Keller, J. (2012). “Flow and its affective, cognitive, and performance-related consequences,” in *Advances in flow research*, ed. S. Engeser (New York, NY: Springer), 65–85. doi: 10.1007/978-1-4614-2359-1_4
- Léger, P. M., Davis, F. D., Cronan, T. P., and Perret, J. (2014). Neurophysiological correlates of cognitive absorption in an enactive training context. *Comput. Hum. Behav.* 34, 273–283. doi: 10.1016/j.chb.2014.02.011
- Luft, C. D. B., and Bhattacharya, J. (2015). Aroused with heart: modulation of heartbeat evoked potential by arousal induction and its oscillatory correlates. *Sci. Rep.* 5, 1–11. doi: 10.1038/srep15717
- Mahinrad, S., Jukema, J. W., Van Heemst, D., MacFarlane, P. W., Clark, E. N., De Craen, A. J. M., et al. (2016). 10-Second heart rate variability and cognitive function in old age. *Neurology* 86, 1120–1127. doi: 10.1212/WNL.0000000000002499
- Malik, M., Bigger, J. T., Camm, A. J., Kleiger, R. E., Malliani, A., Moss, A. J., et al. (1996). Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. *Eur. Heart J.* 17, 354–381. doi: 10.1093/oxfordjournals.eurheartj.a014868
- Matthews, G., Warm, J. S., Reiner, L. E., Langheim, L. K., and Saxby, D. J. (2010). “Task Engagement, Attention, and Executive Control,” in *Handbook of Individual Differences in Cognition*, eds A. Gruszka, G. Matthews, and B. Szymura (Heidelberg: Springer Science + Business Media), 205–230. doi: 10.1007/978-1-4419-1210-7_13
- Montoya, P., Schandry, R., and Müller, A. (1993). Heartbeat evoked potentials (HEP): topography and influence of cardiac awareness and focus of attention. *Electroencephalogr. Clin. Neurophysiol.* 88, 163–172. doi: 10.1016/0168-5597(93)90001-6
- Nacke, L. E., and Lindley, C. A. (2010). Affective Ludology, Flow and Immersion in a First- Person Shooter: measurement of Player Experience. *ArXiv [Preprint]* Available Online at: <http://arxiv.org/abs/1004.0248> (accessed March 2020).
- Narkiewicz, K., van de Borne, P., Montano, N., Hering, D., Kara, T., and Somers, V. K. (2006). Sympathetic Neural Outflow and Chemoreflex Sensitivity Are Related to Spontaneous Breathing Rate in Normal Men. *Hypertension* 47, 51–55. doi: 10.1161/01.HYP.0000197613.47649.02
- Palmer, J. A., Kreutz-Delgado, K., and Makeig, S. (2006). “Super-Gaussian mixture source model for ICA,” in *Independent Component Analysis and Blind Signal Separation. ICA 2006. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* eds J. Rosca, D. Erdogmus, J. C. Principe, and S. Haykin (Berlin, Heidelberg: Springer), 3889, 854–861. doi: 10.1007/11679363_106
- Palmer, J. A., Kreutz-Delgado, K., and Makeig, S. (2011). *AMICA: An adaptive mixture of independent component analyzers with shared components*. San Diego, CA: Swartz Center for Computational Neuroscience.
- Palmer, J. A., Makeig, S., Kreutz-Delgado, K., and Rao, B. D. (2008). “Newton method for the ICA mixture model,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*, (Las Vegas, NV, USA: IEEE), 1805–1808. doi: 10.1109/ICASSP.2008.4517982
- Park, H. D., and Blanke, O. (2019). Heartbeat-evoked cortical responses: underlying mechanisms, functional roles, and methodological considerations. *NeuroImage* 197, 502–511. doi: 10.1016/j.neuroimage.2019.04.081
- Park, H. D., Bernasconi, F., Bello-Ruiz, J., Pfeiffer, C., Salomon, R., and Blanke, O. (2016). Transient modulations of neural responses to heartbeats covary with bodily self-consciousness. *J. Neurosci.* 36, 8453–8460. doi: 10.1523/JNEUROSCI.0311-16.2016
- Park, H. D., Correia, S., Ducorps, A., and Tallon-Baudry, C. (2014). Spontaneous fluctuations in neural responses to heartbeats predict visual detection. *Nat. Neurosci.* 17, 612–618. doi: 10.1038/nn.3671
- Peifer, C., Schulz, A., Schächinger, H., Baumann, N., and Antoni, C. H. (2014). The relation of flow-experience and physiological arousal under stress - Can u shape it? *J. Exp. Soc. Psychol.* 53, 62–69. doi: 10.1016/j.jesp.2014.01.009
- Perone, S., Weybright, E. H., and Anderson, A. J. (2019). Over and over again: changes in frontal EEG asymmetry across a boring task. *Psychophysiology* 56, 1–12. doi: 10.1111/psyp.13427
- Petzschner, F. H., Weber, L. A., Wellstein, K. V., Paolini, G., Do, C. T., and Stephan, K. E. (2019). Focus of attention modulates the heartbeat evoked potential. *NeuroImage* 186, 595–606. doi: 10.1016/j.neuroimage.2018.11.037
- Pfeiffer, C., and De Lucia, M. (2017). Cardio-audio synchronization drives neural surprise response. *Sci. Rep.* 7, 1–10. doi: 10.1038/s41598-017-13861-8
- Pion-Tonachini, L., Kreutz-Delgado, K., and Makeig, S. (2019). ICLabel: an automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage* 198, 181–197. doi: 10.1016/j.neuroimage.2019.05.026
- Pollatos, O., and Schandry, R. (2004). Accuracy of heartbeat perception is reflected in the amplitude of the heartbeat-evoked brain potential. *Psychophysiology* 41, 476–482. doi: 10.1111/1469-8986.2004.00170.x
- Porges, S. W. (1995). Cardiac vagal tone: a physiological index of stress. *Neurosci. Biobehav. Rev.* 19, 225–233. doi: 10.1016/0149-7634(94)00066-A
- Porges, S. W. (2001). The polyvagal theory: phylogenetic substrates of a social nervous system. *Int. J. Psychophysiol.* 42, 123–146. doi: 10.1016/S0167-8760(01)00162-3
- Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proc. Natl. Acad. Sci. U. S. A.* 98, 676–682. doi: 10.1073/pnas.98.2.676
- Rheinberg, F., and Vollmeyer, R. (2003). Flow-Erleben in einem Computerspiel unter experimentell variierten Bedingungen. *Z. Für Psychol.* 4, 161–170. doi: 10.1026/0044-3409.211.4.161
- Rutrecht, H., Wittmann, M., Khoshnoud, S., and Igarzábal, F. A. (2021). Time Speeds Up During Flow States: a Study in Virtual Reality with the Video Game Thumper. *Timing Time Percept.* 9, 353–376. doi: 10.1163/22134468-bja10033
- Sadlo, G. (2016). “Towards a Neurobiological Understanding of Reduced Self-Awareness During Flow: An Occupational Science Perspective,” in *Flow Experience: Empirical Research and Applications*, eds L. Harmat, F. Ø Andersen, F. Ullén, J. Wright, and G. Sadlo (Berlin: Springer), 375–388. doi: 10.1007/978-3-319-28634-1
- Salen, K., and Zimmerman, E. (2003). *Rules of Play: Game Design Fundamentals*. Cambridge, MA: The MIT Press.
- Sassenhagen, J., and Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology* 56:e13335. doi: 10.1111/psyp.13335
- Schandry, R., and Montoya, P. (1996). Event-related brain potentials and the processing of cardiac activity. *Biol. Psychol.* 42, 75–85. doi: 10.1016/0301-0511(95)05147-3
- Schulz, A., Köster, S., Beutel, M. E., Schächinger, H., Vögele, C., Rost, S., et al. (2015). Altered patterns of heartbeat-evoked potentials in depersonalization/derealization disorder: neurophysiological evidence for impaired cortical representation of bodily signals. *Psychosom. Med.* 77, 506–516. doi: 10.1097/PSY.0000000000000195

- Schulz, A., Strelzyk, F., Ferreira de Sá, D. S., Naumann, E., Vögele, C., and Schächinger, H. (2013). Cortisol rapidly affects amplitudes of heartbeat-evoked brain potentials—Implications for the contribution of stress to an altered perception of physical sensations?. *Psychoneuroendocrinology* 38, 2686–2693. doi: 10.1016/j.psyneuen.2013.06.027
- Sel, A., Azevedo, R. T., and Tsakiris, M. (2017). Heartfelt Self: cardio-Visual Integration Affects Self-Face Recognition and Interoceptive Cortical Processing. *Cereb. Cortex* 27, 5144–5155. doi: 10.1093/cercor/bhw296
- Seth, A. K., Suzuki, K., and Critchley, H. D. (2012). An interoceptive predictive coding model of conscious presence. *Front. Psychol.* 3:395. doi: 10.3389/fpsyg.2011.00395
- Shaffer, F., and Ginsberg, J. P. (2017). An Overview of Heart Rate Variability Metrics and Norms. *Front. Public Health* 5:258. doi: 10.3389/fpubh.2017.00258
- Shao, S., Shen, K., Wilder-Smith, E. P. V., and Li, X. (2011). Effect of pain perception on the heartbeat evoked potential. *Clin. Neurophysiol.* 122, 1838–1845. doi: 10.1016/j.clinph.2011.02.014
- Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., and Leahy, R. M. (2011). Brainstorm: a User-Friendly Application for MEG/EEG Analysis. *Comput. Intell. Neurosci.* 2011, 1–13. doi: 10.1155/2011/879716
- Tallon-Baudry, C., Campana, F., Park, H. D., and Babo-Rebelo, M. (2018). The neural monitoring of visceral inputs, rather than attention, accounts for first-person perspective in conscious vision. *Cortex* 102, 139–149. doi: 10.1016/j.cortex.2017.05.019
- Teghil, A., Di Vita, A., D'Antonio, F., and Boccia, M. (2020). Inter-individual differences in resting-state functional connectivity are linked to interval timing in irregular contexts. *Cortex* 128, 254–269. doi: 10.1016/j.cortex.2020.03.021
- Tian, Y., Bian, Y., Han, P., Wang, P., Gao, F., and Chen, Y. (2017). Physiological signal analysis for evaluating flow during playing of computer games of varying difficulty. *Front. Psychol.* 8:1121. doi: 10.3389/fpsyg.2017.01121
- Tobin, S., Bisson, N., and Grondin, S. (2010). An ecological approach to prospective and retrospective timing of long durations: a study involving gamers. *PLoS One* 5:e9271. doi: 10.1371/journal.pone.0009271
- Tsakiris, M., and De Preester, H. (2018). *The Interoceptive Mind: From Homeostasis to Awareness*. Oxford: Oxford University Press.
- Ullén, F., de Manzano, Ö, Theorell, T., and Harmat, L. (2010). “The Physiology of Effortless Attention: Correlates of State Flow and Flow Proneness,” in *Effortless attention: a new perspective in the cognitive science of attention and action*, ed. B. Bruya (Cambridge (Massachusetts): MIT Press), 205–218. doi: 10.7551/mitpress/9780262013840.003.0011
- Ulrich, M., Keller, J., and Grön, G. (2016). Neural signatures of experimentally induced flow experiences identified in a typical fMRI block design with BOLD imaging. *Soc. Cogn. Affect. Neurosci.* 11, 496–507. doi: 10.1093/scan/nsv133
- Ulrich, M., Keller, J., Hoenig, K., Waller, C., and Grön, G. (2014). Neural correlates of experimentally induced flow experiences. *NeuroImage* 86, 194–202. doi: 10.1016/j.neuroimage.2013.08.019
- Ulrich, M., Niemann, J., Boland, M., Kammer, T., Niemann, F., and Grön, G. (2018). The neural correlates of flow experience explored with transcranial direct current stimulation. *Exp. Brain Res.* 236, 3223–3237. doi: 10.1007/s00221-018-5378-0
- Veltman, J. A., and Gaillard, A. W. K. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics* 41, 656–669. doi: 10.1080/001401398186829
- Vogel, D., Krämer, K., Schoofs, T., Kupke, C., and Vogeley, K. (2018). Disturbed Experience of Time in Depression—Evidence from Content Analysis. *Front. Hum. Neurosci.* 12:66. doi: 10.3389/fnhum.2018.00066
- Wei, Y., Ramautar, J. R., Colombo, M. A., Stoffers, D., Gómez-Herrero, G., Van Der Meijden, W. P., et al. (2016). I keep a close watch on this heart of mine: increased interoception in Insomnia. *Sleep* 39, 2113–2124. doi: 10.5665/sleep.6308
- Wientjes, C. J. E. (1992). Respiration in psychophysiology: methods and applications. *Biol. Psychol.* 34, 179–203. doi: 10.1016/0301-0511(92)90015-M
- Witowska, J., Schmidt, S., and Wittmann, M. (2020). What happens while waiting? How self-regulation affects boredom and subjective time during a real waiting situation. *Acta Psychol.* 205:103061. doi: 10.1016/j.actpsy.2020.103061
- Wittmann, M. (2013). The inner sense of time: how the brain creates a representation of duration. *Nat. Rev. Neurosci.* 14, 217–223. doi: 10.1038/nrn3452
- Wittmann, M. (2015). Modulations of the experience of self and time. *Conscious. Cogn.* 38, 172–181. doi: 10.1016/j.concog.2015.06.008
- Wittmann, M. (2018). *Altered States of Consciousness: Experiences Out of Time and Self*. Cambridge, MA: MIT Press.
- Wittmann, M., Leland, D. S., Churan, J., and Paulus, M. P. (2007). Impaired time perception and motor timing in stimulant-dependent subjects. *Drug Alcohol Depend.* 90, 183–192. doi: 10.1016/j.drugalcdep.2007.03.005
- Yelamanchili, T. (2018). *Neural Correlates of Flow, Boredom, and Anxiety in Gaming: An Electroencephalogram Study*. Ph.D. thesis. Missouri: Missouri University of Science and Technology.
- Yun, K., Doh, S., Carrus, E., Wu, D., and Shimojo, S. (2017). Being in the zone: flow state and the underlying neural dynamics in video game playing. *ArXiv [Preprint]* Available Online at: <https://arxiv.org/abs/1711.06967> (accessed March 2020).

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.

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.

Copyright © 2022 Khoshnoud, Alvarez Igarzábal and Wittmann. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Music, Math, and Working Memory: Magnetoencephalography Mapping of Brain Activation in Musicians

Ching-I Lu^{1*}, Margaret Greenwald^{1,2}, Yung-Yang Lin^{3,4*} and Susan M. Bowyer^{2,5,6}

¹ Department of Communication Sciences and Disorders, Wayne State University, Detroit, MI, United States, ² Department of Neurology, Wayne State University, Detroit, MI, United States, ³ Institute of Brain Science and Institute of Clinical Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan, ⁴ Department of Critical Care Medicine, Taipei Veterans General Hospital, Taipei, Taiwan, ⁵ Department of Neurology, Henry Ford Health System, Detroit, MI, United States, ⁶ Department of Physics, Oakland University, Rochester, MI, United States

OPEN ACCESS

Edited by:

Vasil Kolev,
Institute of Neurobiology, Bulgarian
Academy of Sciences (BAS), Bulgaria

Reviewed by:

Nadia Justel,
Interdisciplinary Laboratory
of Cognitive Neuroscience, Argentina
Elizabeth W. Pang,
University of Toronto, Canada

*Correspondence:

Ching-I Lu
chingilu@gmail.com
Yung-Yang Lin
g2000kev@gmail.com

Specialty section:

This article was submitted to
Brain Imaging and Stimulation,
a section of the journal
Frontiers in Human Neuroscience

Received: 31 January 2022

Accepted: 25 March 2022

Published: 16 May 2022

Citation:

Lu C-I, Greenwald M, Lin Y-Y and
Bowyer SM (2022) Music, Math,
and Working Memory:
Magnetoencephalography Mapping
of Brain Activation in Musicians.
Front. Hum. Neurosci. 16:866256.
doi: 10.3389/fnhum.2022.866256

Musical transposing is highly demanding of working memory, as it involves mentally converting notes from one musical key (i.e., pitch scale) to another key for singing or instrumental performance. Because musical transposing involves mental adjustment of notes up or down by a specific amount, it may share cognitive elements with arithmetical operations of addition and subtraction. We compared brain activity during high and low working memory load conditions of musical transposing versus math calculations in classically trained musicians. Magnetoencephalography (MEG) was sensitive to differences of task and working memory load. Frontal-occipital connections were highly active during transposing, but not during math calculations. Right motor and premotor regions were highly active in the more difficult condition of the transposing task. Multiple frontal lobe regions were highly active across tasks, including the left medial frontal area during both transposing and calculation tasks but the right medial frontal area only during calculations. In the more difficult calculation condition, right temporal regions were highly active. In coherence analyses and neural synchrony analyses, several similarities were seen across calculation tasks; however, latency analyses were sensitive to differences in task complexity across the calculation tasks due to the high temporal resolution of MEG. MEG can be used to examine musical cognition and the neural consequences of music training. Further systematic study of brain activity during high versus low memory load conditions of music and other cognitive tasks is needed to illuminate the neural bases of enhanced working memory ability in musicians as compared to non-musicians.

Keywords: working memory, musical transposing, calculation, music training, magnetoencephalography (MEG)

INTRODUCTION

Working memory is enhanced in musicians as compared to non-musicians (George and Coch, 2011), but neuroimaging studies of musicians have yielded little information about their brain activity during cognitive tasks with high versus low working memory load conditions. Similarly, there is insufficient detail about high versus low working memory demands of specific music tasks in the general debate about potential cognitive effects of music training (Nutley et al., 2014; Swaminathan et al., 2017).

Working memory is highly taxed in some music tasks; in others, the working memory demand is very low. If music training includes only low demands on working memory, then working memory is not likely to improve from the training, nor would it be expected to influence working memory function that supports a different cognitive behavior, such as math or reading. One music task that is demanding of working memory is musical transposing, which involves mentally converting notes from one musical key (i.e., pitch scale) to another key for singing or instrumental performance. Working memory demands are high during musical transposing because the target musical key must be stored temporarily while notes are manipulated. No studies of the cognitive effects of musical training have included transposing in the music training program, though this would be one means to assess whether music training involving high working memory load would improve working memory capacity.

Ongoing research into the cognitive and neural effects of training in music includes studies of experts and non-experts, and the longitudinal effects of exposure and training in children (Schön et al., 2002; Stewart et al., 2003). Music and math are separate cognitive domains, and separable from language (Ivanova et al., 2020), although there is some evidence they may share domain-general structural processing mechanisms with language (Van de Cavey and Hartsuiker, 2016; Nakai and Okanoya, 2018). There is evidence for a relationship between musical achievement and math achievement, usually associative (Holochwost et al., 2017) rather than causal (Hille and Schupp, 2015; Wallick, 1998). Recently, Bergee and Weingarten (2021) controlled for background variables that may influence achievement in music, math, and reading in children, and found that musical achievement did relate to math and reading achievement. However, following a meta-analytic review of studies of music training, Sala and Gobet (2020) concluded that music training has no impact on non-music cognitive skills and academic achievement. In studies of the effects of music training, greater specificity is needed in descriptions of the cognitive components that are highly active during the music training tasks, and whether the training tasks involve cognitive abilities from domains other than music.

Math, for example, is linked to the musical transposing task in that changing from one musical key to another is based on mental calculations that can involve addition or subtraction skills. A varying potential for unidirectional or bidirectional influence of learning in music and math tasks will depend on the cognitive components needed to accomplish each task, and the type and degree of overlap between these cognitive components across tasks. Transposing is one example of a music task in which the influences of music and math during training of transposing may be bidirectional.

In previous reports, we have shown that magnetoencephalography (MEG) is sensitive to differences in working memory load (Lu et al., 2019, 2021). By comparing brain activity in musical transposing of musical notation versus sight-reading (in which notes are played as written), we observed that the additional mental conversion required for transposing was linked to slowed activation of the ventral (fusiform gyrus)

occipito-temporal stream of visual-spatial encoding, and to increased frontal lobe activation (Lu et al., 2021).

Further studies are needed to compare aspects of music, language and mathematical cognition, and brain activity that supports them in musicians and non-musicians. For example, studies using visual music and math tasks are needed to determine the relative roles of domain-general skills in working memory and spatial attention (Sligte et al., 2009) compared to domain-specific cognitive abilities in music or math.

Because musical transposing involves mental calculations for modifying musical keys, it may share cognitive elements with arithmetical operations of addition and subtraction. It is therefore of interest to compare the brain activity underlying simple arithmetic to that underlying the mental conversion of transposing in which notes are adjusted up or down by a specific amount. Further, imposing a requirement that participants hold a cue in working memory to perform the arithmetic and transposing tasks adds another level of similarity across the tasks and a way for working memory load to be manipulated up or down. By studying similarities and differences across the arithmetic and transposing tasks of higher and lower working memory load, we obtain clues as to how these cognitive tasks are related and how training in one (e.g., music training) could possibly affect the other (e.g., performance in math).

In the current paper, we report brain activity in trained musicians during math calculation tasks compared to musical transposing using MEG. We also examine the effects of working memory load across music and calculation tasks.

MATERIALS AND METHODS

Participants

Twenty-one participants at Taipei Veterans General Hospital, Taipei, Taiwan who were able to read the Western musical notation system completed the study voluntarily with normal vision, hearing, motor and cognitive abilities. All of the data from one participant had to be discarded due to sleeping, and individual transposing and calculation task data from two participants had to be discarded due to noise; thus, data on all tasks were analyzed for 19 participants. All participants were female classically trained musicians (age range = 19–28; $x = 23.71$) with at least 10 years of musical instrumental training (range = 10–22 years; $x = 16.71$), including reading of standard musical notation. Potential participants were not included if their major instrument was a transposing instrument (i.e., an instrument that produces a higher or lower pitch than is shown in music notation written for it). For example, if a musician's major instrument is clarinet in B flat, they would always automatically lower two interval pitches while they read the score in G clef. Thus, long-term intensive experience in transposing in these individuals has the potential to alter patterns of brain activity during transposing as compared to other musicians. For the musicians who were included in this study, a transposing instrument was not the major instrument and, based on typical music training in Taiwan in which students are trained to read transposing scores from approximately the 3rd to the 12th

grade, the average amount of experience in transposing for the participants included in this study was approximately 10 years.

All participants completed the Edinburgh Handedness Inventory (Oldfield, 1971), and laterality quotients indicated strong right-handed preference for all but one participant, who was ambidextrous. They also passed the Mini-Mental State Examination (MMSE; Folstein et al., 1975), which screened for cognitive impairment, and the digit span task (WAIS-III; Wechsler, 1997), which measured working memory storage capacity. All participants were within normal range. None had a history of neurological or psychiatric diseases, or developmental learning difficulties. Each participant provided a written informed consent prior the experiment. The study was approved by the Taipei Veterans General Hospital Human Research Review Board, Taipei.

Experimental Design and Stimuli

A two-factor within-participant design was used: 2_{within} (stimuli: Musical notation versus Digits) \times 2_{within} (task: Easier – one note or one single-digit versus Difficult-five notes or five-single digits), with location, amplitude, and latency of activation as dependent variables. Brain activation was observed during the four experimental tasks below. We previously discussed the results of the transposing tasks in comparison to musical sight-reading (Lu et al., 2021).

One Single-Digit Calculation (1D)

A written cue indicated a plus (+) or minus (–) single digit from 1 to 5 presented for 1,000 ms randomly (+ 1, + 2, + 3, + 4, + 5, – 1, – 2, – 3, – 4, – 5). After a 1,000 ms blank screen, each stimulus ($n = 60$) was presented randomly for 1,500 ms plus 1,000 ms ISI. Each stimulus was a single digit from 0 to 9. During each presentation, the participant was to silently add to or subtract from the target stimulus based on the plus or minus cue and then silently name the correct answer, with no overt action (e.g., the cue + 4 followed by the target 8 would be silently named as 12). This task lasted 4 min 30 s.

Five Single-Digit Calculation (5D)

This task was identical to the one single-digit calculation task except that a sequence of five digits was presented for 3,500 ms (Figure 1). During each presentation, the participant was to silently add to or subtract from the target stimulus based on the plus or minus cue and then silently name the correct answer, with no overt action (e.g., the cue + 4 followed by the target 6 4 2 8 5 would be silently named, sequentially, as 10, 8, 6, 12, 9). This task lasted 6 min 30 s.

One-Note Transposing With Treble Clef (1T)

A written cue of one of five transposing instruments was presented for 1,000 ms pseudo randomly (Clarinet in A, in E^b or in B^b, French horn in F, and Trumpet in B^b). After 1,000 ms blank screen, each stimulus ($n = 60$) was presented randomly for 1,500 ms plus 1,000 ms ISI. During each presentation, the participant was to silently transpose from the written note to the target key and then silently name the new note, with no overt action. This task lasted 4 min 30 s.

Five-Note Transposing With Treble Clef (5T)

This task was identical to the 1T task except that a sequence of five notes was presented for 3,500 ms (Figure 1). During each presentation, the participant was to silently transpose each written note to the target key and then silently name the new notes sequentially, with no overt action. This task lasted 6 min 30 s.

Procedure

Stimuli were presented electronically using E-Prime Professional 2.0 software (Psychology Software Tools, Pittsburgh, PA). To document participant accuracy in the tasks, behavioral practice data were collected using overt naming before each participant entered the MEG scanner; over 60 trials of each task, the average pre-test single accuracy was 99% for 1D, 95% for 5D, 87% for 1T, and 80% for 5T. Silent naming was required inside the scanner due to decreased signal noise (due to mouth movement) in silent naming compared to overt naming. Participant brain waves were monitored during tasks in the scanner to ensure participant alertness.

Inside the scanner, oral and written instructions were given immediately prior to each task, with 2-min breaks between tasks. The 1T task was given before the 5T task followed by the 1D and 5D tasks. Including screening, practice and experimental tasks, the procedure lasted approximately 90 min (Figure 1). Immediately following the MEG scan, each participant was asked a general open-ended question as to how they completed the transposing and math tasks.

Data Acquisition and Preprocessing

A 306-channel MEG system (Vectorview, Elekta-Neuromag, Helsinki, Finland) was used; this helmet-shaped device covers the entire adult head except for the face. Participants were monitored continuously by intercom and camera. During data collection, participants were asked to avoid eye and body movements. MEG data were recorded with a high pass filter of 0.1 Hz, low pass filter of 100 Hz, and sampling rate of 508.63 Hz.

MEG signals measured the magnetic fields produced by currents fed into four head position indicator coils at known scalp locations, two high behind the earlobes and two wide apart high on the forehead. Coil locations were chosen in relation to three anatomical landmarks, including left preauricular point, right preauricular point, and nasion, which were determined with three-dimensional digitization. The individual sensors were magnetometers. Head shape was digitized for coregistration to the standard female brain template of T1-weighted MRI. The MRI scan was performed on a GE 1.5-T, 1-m-bore whole body magnet. MRI scan parameters were coronal T1 images, 124 slices, and 256×256 matrix including the entire skin surface of the head. A model of cortical brain surface was created from this standard MRI and performed in MEG-TOOLS (Moran et al., 2005). The MRI was segmented and brain surface was represented by a cortical model of approximately 4,000 dipoles each having x, y, and z orientation at each site. Sites were distributed to represent the same volume of cortical gray matter. This model was then morphed to fit the digitized head shape collected during MEG acquisition.


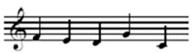
A	Instruction	-3	Blank	5	Blank
	Trigger	1000 ms	1000 ms	1500 ms	1000ms
B	Instruction	+2	Blank	6 1 5 2 4	Blank
	Trigger	1000 ms	1000 ms	3500 ms	1000 ms
C	Instruction	Clarinet in B^b	Blank		Blank
	Trigger	1000 ms	1000 ms	1500 ms	1000ms
D	Instruction	Clarinet in A	Blank		Blank
	Trigger	1000 ms	1000 ms	3500 ms	1000 ms

FIGURE 1 | Sample stimuli, duration and procedure across the four experimental tasks: **(A)** the one single-digit calculation task (1D); **(B)** the five single-digit calculation task (5D); **(C)** the one-note transposing task (1T); and **(D)** the five-note transposing task (5T). All stimuli for 1T and 5T were presented on treble clef and all cues (transposing instruments) were presented using word form.

Using an independent component analysis (ICA), noise artifacts due to heart and body movement were eliminated in post-processing. Any other artifacts in the data were removed if needed using singular valued decomposition. Regarding movement artifact, runs would have been repeated if the coil on head positions exceeded 0.5 cm, although this did not occur during data acquisition. Data were filtered 3–85 Hz with notch at 60 Hz. The locations of events on trigger and response channels were used to select 1.5-s epochs of MEG data to examine average evoked responses during the four experimental tasks.

MR-FOCUSS

Event-related cortical activation was studied by averaging all 60 trials of the participant's measured evoked MEG field responses during each task. Data were analyzed using MR-FOCUSS (current distribution technique; Moran et al., 2005) to localize and quantify cortical activation within the brain. The latency (in ms), location and average amplitude of response (nAm/time point) were extracted from MR-FOCUSS imaging results. MR-FOCUSS cortical mapping was applied to the interval 0–1,500 ms after stimulus onset in each experimental task. Selection of significant cortical activation was determined by visually inspecting imaged MEG solutions overlaid on the anatomical MRI and setting the display threshold to 30% (color coded blue) of the maximum cortical source amplitude (color coded red), and by selecting the high peaks of activity relative to the background brain noise.

Coherence Source Imaging

Synchronization of neuronal activity was quantified by calculating coherence between cortical sites from MEG imaged brain activation (Elisevich et al., 2011; Bowyer, 2016). A model of

the cortical brain surface was created from an age- appropriate standard MRI of a female brain, as described above. To calculate coherence, the MEG data were first divided into 40 (1T or 1D tasks) or 52 (5T or 5D tasks) segments each containing 7.5-s segments of data, and cortical activity in each segment was imaged on the MRI using the MR-FOCUSS imaging technique. Using the time sequence of imaged activity, coherence between active cortical model sites was calculated for each data segment and then averaged for the completed study. In addition, for each cortical model site, connectivity was quantified by a histogram of the number of sites to which the site had the same level of coherence. Statistical analysis of cortical coherence levels (0 to 1) were used to quantify differences in network connectivity between groups. Changes in coherence and connectivity between brain regions implicated as having deviant electrophysiological activity in different tasks within the participants' brains were quantified and included in further statistical analysis.

A region-of-interest (ROI) tool implemented in MEG Tools was used to identify 54 regions in the brain (27 in each hemisphere). MEG Tools uses a non-linear volumetric transformation of the participant's brain to transform MEG coordinates to standard Talairach or MNI coordinates. This enables the ROI tool to access an atlas of Brodmann's area identifiers and an atlas of cortical structures.

Neural Synchrony

T-test was used to assess task difference in average coherence values for each pair of brain regions ($N = 1431$) (Lajiness-O'Neill et al., 2014). A p value was produced for each region pair. Because of the large number of tests being performed simultaneously, using a significance level of $\alpha = 0.1$ without adjusting for multiple testing would lead to a large number of

false positive results; therefore, false discovery (FDR) was used to adjust for multiple testing. Bonferroni adjustments for multiple comparisons aim to control the family wise error rate. From each *t*-test, a *t*-score was computed according to the method of Efron to summarize the difference in coherence values between tasks.

RESULTS

Latency and Amplitude

Cortical mapping using MR-FOCUSS analysis displayed multiple areas of neuronal activity, including visual cortex, fusiform gyrus, superior temporal gyrus (STG), angular gyrus (AG), supramarginal gyrus (SMG; Wernicke's area included activation of the STG, AG, and SMG), the superior parietal gyrus, and frontal lobe regions. Simultaneous activation of both visual and frontal gyri was also measured. Selected images from an individual participant are shown as examples in **Figures 2, 3**. Note the latencies for this individual fall within the midrange of the average across all subjects (showing simultaneous frontal and occipital activation occurring earlier for the 5D task than for the 5T task). The temporal resolution of peak activation in these areas during the four tasks is summarized in **Table 1**.

To better understand the three main effects [ROI (region of interest), task, memory load and their interaction effects], a three-factor repeated-measure analysis was conducted 5_{within} (ROI: visual versus fusiform versus Wernicke's area versus superior parietal versus visual + frontal areas) $\times 2_{\text{within}}$ (task: calculation versus transposing) $\times 2_{\text{within}}$ (memory loading: one versus five) with peak latency of activity (visual, fusiform, superior parietal, Wernicke's area, and visual + frontal) as dependent

variables. Using general liner model – repeated measure revealed significant main effects of ROI [$F(4,68) = 360.24, p < 0.001$], task [$F(1,17) = 51.00, p < 0.001$], and loading [$F(1,17) = 6.45, p = 0.021$]. There is an interaction effect between ROI and task [$F(4,68) = 67.84, p < 0.001$]. Main effects of ROI showed no significant difference in early latency of activity across the four tasks (including visual, fusiform, and superior parietal). However, the four tasks showed significantly different latency at Wernicke's area [$F(3,72) = 3.64, p = 0.017$] and at the stage of simultaneous visual and frontal activation [$F(3,72) = 56.35, p < 0.001$]. *Post hoc* comparisons of 1D versus 1T showed a significant difference at Wernicke's area [$t(18) = 3.47, p = 0.003$] such that 1T (199 ± 68 ms) was faster than 1D (274 ± 99 ms); for the 1T versus 5T comparison [$t(17) = -3.38, p = 0.004$], 1T (202 ± 69 ms) was faster than 5T (272 ± 61 ms). At the stage of simultaneous visual and frontal activation, a significant difference was indicated for the 1D versus 1T comparison [$t(18) = -10.12, p < 0.001$] such that 1D (548 ± 103 ms) was faster than 1T (968 ± 176 ms), for the 1D versus 5D comparison [$t(18) = -3.05, p = 0.007$] such that 1D (548 ± 103 ms) was faster than 5D (666 ± 113 ms), and for the 5D versus 5T comparison [$t(17) = -8.62, p < 0.001$] such that 5D (668 ± 117 ms) was faster than 5T (1028 ± 125 ms).

Coherence Source Imaging

We analyzed the entire dataset for each task to find the top five highest coherence regions active for all of the participants combined (**Table 2**). During 1D, 5D, and 1T tasks, the highest coherent region was left parahippocampus; in 5T, the highest coherent region was right precentral motor cortex. Interestingly, the transposing tasks and calculation tasks all engaged the left

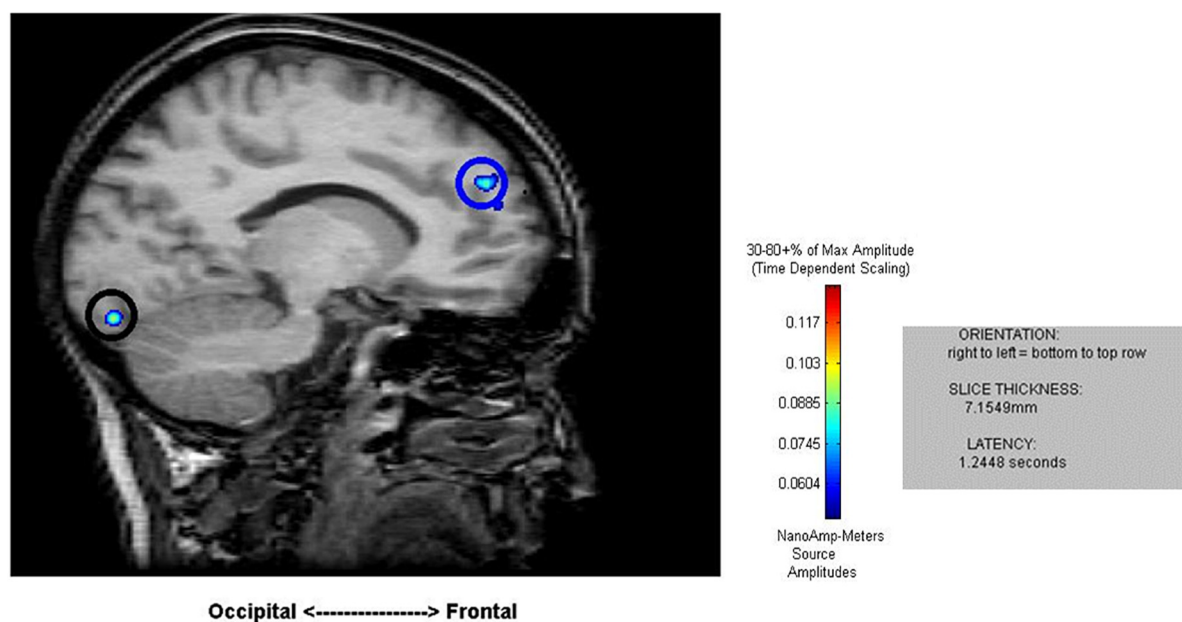


FIGURE 2 | Sample results of the MEG recordings (MR-FOCUSS analyses) for an individual participant showing simultaneous frontal and occipital activation at 1,244 ms latency in five-note transposing (5T).

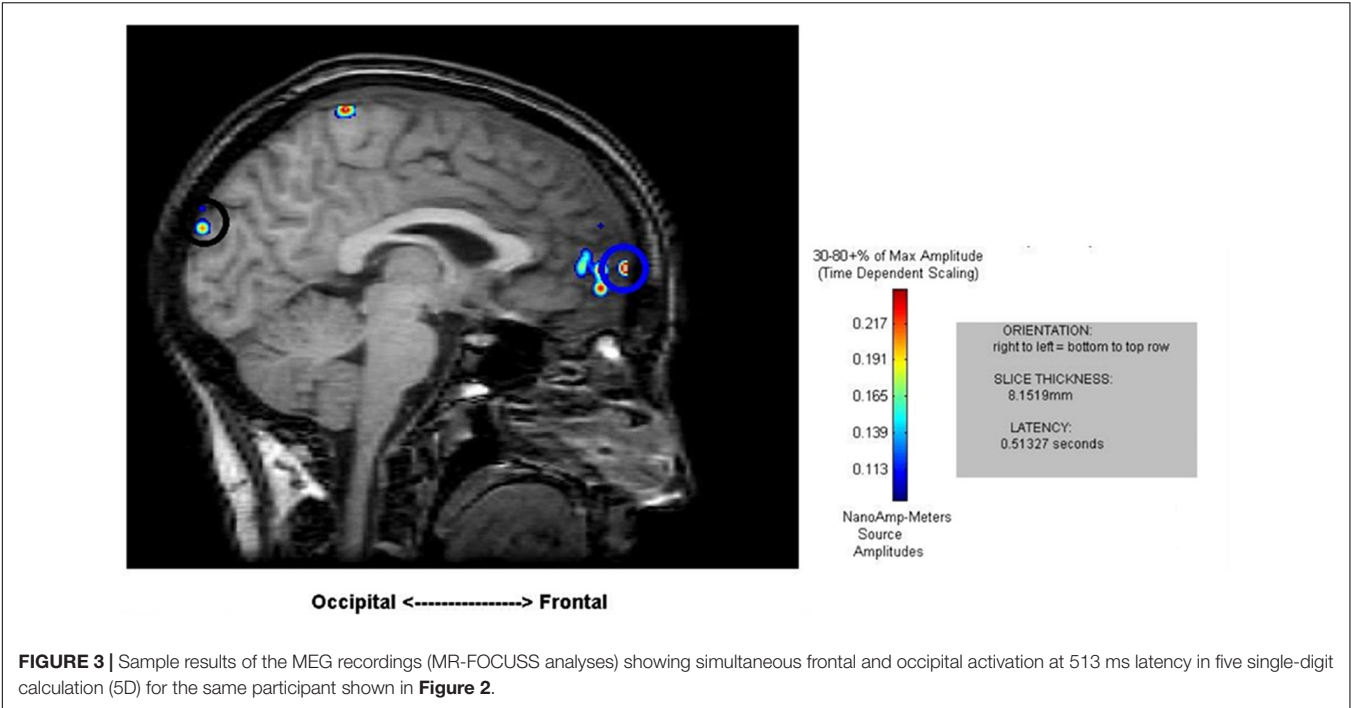


TABLE 1 | Temporal resolution of magnetoencephalography (MEG) signals arising from the peak brain activation during each task

	Visual cortex	Fusiform gyrus	Superior parietal gyrus	Wernicke's area ^a	Frontal and visual cortex ^b
One single-digit calculation (1D)	84 ± 14 ms	244 ± 131 ms	241 ± 107 ms	274 ± 99 ms	548 ± 103 ms
One-note transposing (1T)	83 ± 20 ms	237 ± 144 ms	216 ± 72 ms	199 ± 68 ms	968 ± 176 ms
Five single-digit calculation (5D)	91 ± 14 ms	236 ± 142 ms	257 ± 91 ms	282 ± 111 ms	666 ± 113 ms
Five-note transposing (5T)	84 ± 16 ms	241 ± 129 ms	270 ± 122 ms	271 ± 60 ms	1019 ± 127 ms

^aBilateral superior temporal gyrus (STG), angular gyrus (AG), and supramarginal gyrus (SMG).

^bFrontal and visual regions activated simultaneously.
ms, milliseconds.

TABLE 2 | Spatial resolution of magnetoencephalography (MEG) signals arising from the top five highest coherent regions during each task.

	The highest region	2nd	3rd	4th	5th
One single-digit calculation (1D)	Left parahippocampus	Right medial frontal	Left medial frontal	Left inferior frontal	Right superior frontal
One-note transposing (1T)	Left parahippocampus	Left superior parietal	Right superior frontal	Right medial orbitofrontal	Left medial frontal
Five single-digit calculation (5D)	Left parahippocampus	Right medial frontal	Left medial frontal	Right middle temporal	Right fusiform
Five-note transposing (5T)	Right precentral (BA4)	Right superior occipital	Right inferior frontal	Right precentral (BA6)	Left medial frontal

medial frontal area. Also, the calculation tasks had the same top three highest coherent regions: Left parahippocampus, right medial frontal, and then left medial frontal.

Neural Synchrony Analysis

To identify neuronal networks most strongly activated during each task, 1,431 pathway connections were evaluated for their coherence value in each task, and then a bootstrap method was used to identify differences ($p < 0.05$) between tasks. This provides information on which networks (i.e., two locations) are significantly involved for each task compared to the other task.

In comparing the 5D networks versus the 5T networks, significant differences were found in 13 out of 1,431 pathways.

In particular, inter-hemispheric network activity differences were identified between limbic system and other regions. The p values of less than 0.05 identified using this procedure are listed in **Table 3**.

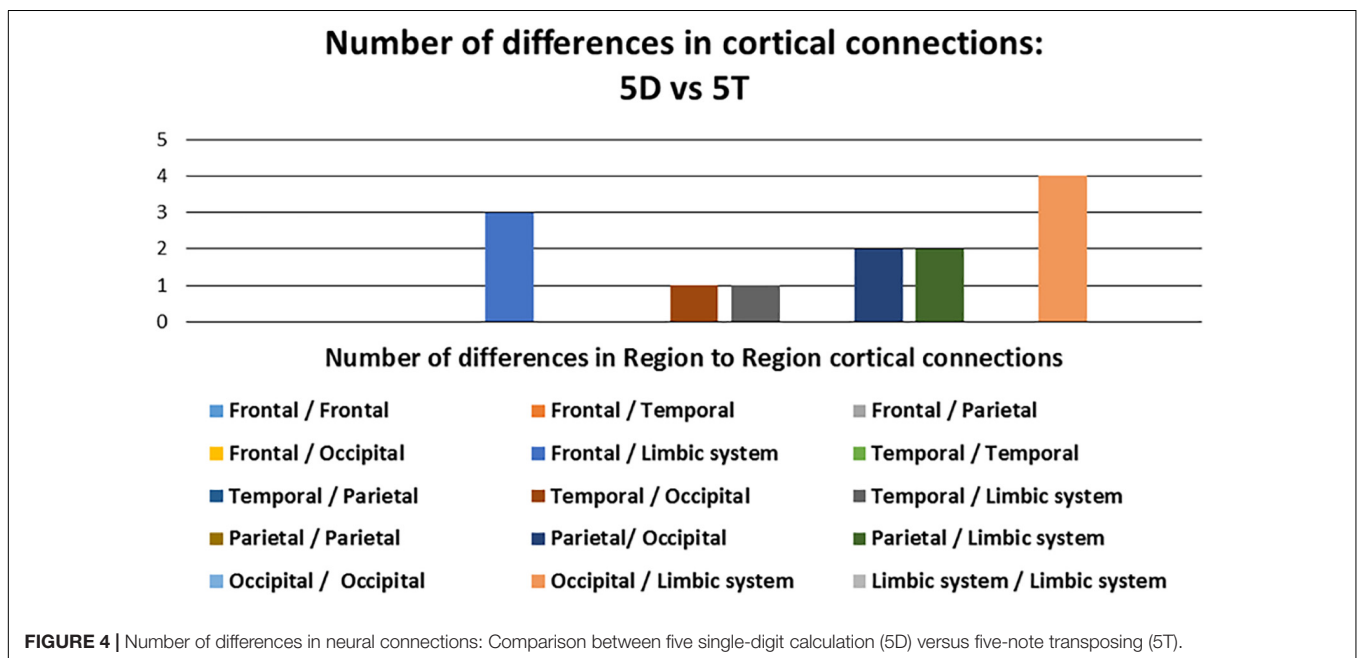
Of the 13 pathways that were found to have significant differences between the 5D networks versus the 5T networks, the most likely (i.e., most common) was the occipital-limbic system pathway. This is shown in **Figure 4**, as compared to the less likely pathways.

In comparing the 1D networks versus the 1T networks, significant differences were found in 3 out of 1,431 pathways. In particular, inter-hemispheric network activity differences were identified between occipital and frontal regions.

TABLE 3 | Differences in network activation: Five single-digit calculation (5D) versus five-note transposing (5T).

Path	Mean.5D	Mean.5T	t	p.value
L.cingulate_gyrus.L.inferior_frontal_gyrus	0.027	0.021	2.109	0.042
R.cingulate_gyrus.R.gyrus_rectus	0.03	0.02	2.057	0.047
L.inferior_occipital_gyrus.R.parahippocampal_gyrus	0.053	0.073	-2.586	0.014
L.inferior_occipital_gyrus.R.insular_cortex	0.028	0.04	-2.265	0.03
R.parahippocampal_gyrus.R.supramarginal_gyrus	0.037	0.049	-2.192	0.035
L.inferior_occipital_gyrus.R.angular_gyrus	0.163	0.197	-2.18	0.036
L.inferior_occipital_gyrus.R.superior_temporal_gyrus	0.135	0.167	-2.131	0.04
R.insular_cortex.R.lingual_gyrus	0.014	0.02	-2.128	0.04
L.lateral_orbitofrontal_gyrus.R.insular_cortex	0.022	0.03	-2.098	0.043
L.inferior_occipital_gyrus.R.supramarginal_gyrus	0.155	0.187	-2.084	0.044
R.angular_gyrus.R.parahippocampal_gyrus	0.038	0.05	-2.084	0.044
L.lingual_gyrus.R.insular_cortex	0.015	0.021	-2.039	0.049
L.middle_temporal_gyrus.R.insular_cortex	0.024	0.031	-2.027	0.05

The limbic system including insula, putamen, parahippocampal gyrus, caudate, hippocampus.



Of the 3 pathways that were found to have significant differences between the 1D networks versus the 1T networks, the most likely (i.e., most common) was the frontal-occipital pathway, especially in right inferior frontal gyrus. The negative *t* values showed that the 1T networks are more active than 1D networks between the occipital and frontal regions (Table 4).

The 1D networks versus the 5D networks were compared. Significant differences were found in none of 1,431 pathways.

In neural synchrony analysis, 2 of the 1,431 pathways differed significantly between the 5T versus 1T. Strong network differences between these transposing tasks were observed in connections from left superior occipital gyrus to left frontal regions (Table 5). As indicated by the positive *t* values, the 5T task involved these pathways significantly more than 1T.

We also observed the following intra- and inter-hemispheric differences in coherence across tasks of calculation versus

transposing of high and low working memory load (5D versus 5T, 1D versus 1T; Figure 5). Of the pathways found to have significant differences between the two tasks of high working memory load (5D versus 5T), the most likely pathways were inter-hemispheric.

After the MEG session, each participant responded to an informal query about their strategy in the experimental tasks. The participants reported that they responded one-by-one to the five digits or notes presented in each 5D and 5T stimulus; further information about individual strategy was not obtained.

DISCUSSION

MEG is sensitive to differences in brain activation in musical transposing versus digit calculation, as shown in this study, and

TABLE 4 | Differences in network activation: One single-digit calculation (1D) versus one note transposing (1T).

Path	Mean.1D	Mean.1T	T	p.value
L.inferior_occipital_gyrus.	0.185	0.22	-2.244	0.031
R.inferior_frontal_gyrus				
L.superior_occipital_gyrus.	0.133	0.163	-2.074	0.045
R.inferior_frontal_gyrus				
L.middle_occipital_gyrus.	0.178	0.21	-2.062	0.046
R.inferior_frontal_gyrus				

TABLE 5 | Differences in network activation: Five-note transposing (5T) versus one-note transposing (1T).

Path	Mean.5T	Mean.1T	T	p.value
L.inferior_frontal_gyrus.	0.147	0.121	2.877	0.007
L.superior_occipital_gyrus				
L.precentral_gyrus.	0.17	0.147	2.025	0.05
L.superior_occipital_gyrus				

can be used to examine the neural correlates of musical and mathematical cognition and the consequences of music training. The patterns of brain activation observed here are influenced by working memory load and task type.

Transposing differed from calculation in frontal-occipital activation. The simultaneous frontal and occipital activation occurred significantly more slowly during transposing compared to calculation (1T slower than 1D; 5T slower than 5D). Neural synchrony analyses revealed more frontal-occipital neural connections active in 5T than 5D, and in 1T than in 1D. Frontal-occipital interactions support visual perception (Ruff et al., 2006) and visual working memory (Barton and Brewer, 2013). In the current study, the notes to be transposed are perceived among

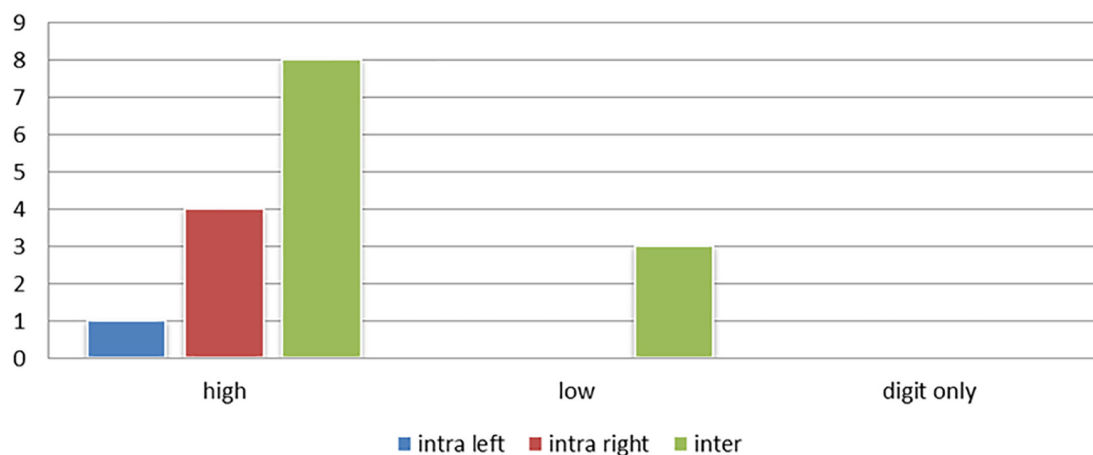
multiple elements that indicate clef, staff, and key signature, whereas in calculation the viewed digits form a simpler display. In transposing, high demands on visual working memory and visual perception may have required greater support from visual cortex and its interactions with frontal lobe systems.

The 5T task involved left occipital-frontal network pathways significantly more than the 1T, and the top four most active regions in 5T were in frontal or occipital areas. Greater visual complexity of the visual pre-cue cannot account for increased frontal-occipital activation in the transposing tasks versus calculation, because previously we identified high activity in frontal-occipital areas in both high and low memory load conditions of musical sight-reading wherein there is no pre-cue (Lu et al., 2021). Musical sight-reading stimuli were similar in complexity to transposing stimuli; as in the 5T task, the high- and low-load sight-reading tasks resulted in high activity in frontal lobe areas and right superior occipital gyrus. In contrast, occipital cortex and frontal-occipital connections were not highly activated in the calculation task.

Interestingly, among the top three most highly active areas in 5T were right primary motor cortex (BA4) and right premotor cortex and supplementary motor area (BA6), possibly reflecting motor planning or programming of hand movement (consciously or unconsciously) when musical notation was viewed. Right precentral cortex was also found to be highly active in both high- and low-load sight-reading (Lu et al., 2021). High activity in the motor cortex was not evident during calculation, suggesting the digit stimuli did not result in initiation of motor programming.

In both the coherence analyses (in top three most highly active regions) and the neural synchrony analyses, the calculation tasks were found to have similar patterns of brain activation. However, in latency analysis there was a significantly slower

Number of differences in intra and interhemispheric cortical connections

**FIGURE 5 |** Number of differences in neural connections: Intra-right hemisphere, intra-left hemisphere and inter-hemispheric cortical differences in tasks of high working memory load (5D versus 5T), low working memory load (1D versus 1T), and digits only (5D versus 1D).

peak of simultaneous frontal and occipital activation in 5D compared to 1D. This likely reflects increased working memory demands in 5D compared to 1D, detected by the high temporal resolution of MEG.

Parahippocampal gyrus, part of medial temporal lobe and the limbic system, was the most highly active brain area for the calculation tasks and 1T. This area is involved in visuospatial processing and other cognitive functions such as memory (Aminoff et al., 2013). In neural synchrony comparisons, the 5D task involved the occipital-limbic pathway significantly more than did 5T. Limbic activity was not as prominent in the 5T task, which had high activations in other regions such as right motor and premotor cortex.

Multiple distributed processes and brain regions have been described as contributing to working memory; for example, mental representations encoding visual feature information in temporal-occipital cortex (ventral visual pathway), spatial information in frontal-parietal cortex (dorsal visual pathway), and information such as behavioral significance in frontal cortex (Eriksson et al., 2015; Ren et al., 2019). The right medial temporal lobe provides important support for memory (Jeneson and Squire, 2012), and the fusiform gyrus is involved in higher visual perception and memory (Weiner et al., 2017). In the current study, when the calculation task involved more numbers to be added or subtracted the participants relied more on right medial temporal lobe and right fusiform gyrus, both highly active in 5D but not in 1D. Previously, we (Lu et al., 2021) attributed slowed fusiform activation in transposing compared to musical sight-reading to the additional mental conversion required for transposing; this may require more ventral stream than dorsal stream processing. Interestingly, in the current study with all four tasks more highly demanding of working memory than the musical sight-reading task, there is no significant difference in time course of fusiform activation.

Superior parietal cortex is important for spatial encoding (Stewart, 2005), and was highly active in the left hemisphere during 1T in the current study. As part of a dorsal “where” stream of spatial encoding, this region is involved in encoding stimulus location, whereas a ventral “what” stream that includes the fusiform appears to encode features for stimulus identification (Mishkin and Ungerleider, 1982; Goodale and Milner, 1992). Reading music may involve interaction between the dorsal and ventral streams (Mongelli et al., 2017). Previously, we (Lu et al., 2019) identified bilateral activation of superior parietal cortex during silent reading of English letters and musical notes. In the current study this region likely played a role across tasks but was not prominently activated in 5T, 5D and 1D.

The left medial frontal area was highly active during all four experimental tasks; interestingly, right medial frontal area was highly active during calculation tasks but not during transposing tasks. Functions of frontal lobe regions may overlap in supporting working memory (Duncan and Owen, 2000); within this distribution of function the right medial frontal area may have some relative specialization for supporting calculation. Differences in working memory demands across tasks have influenced efforts to define frontal lobe specialization in mathematical tasks; for example, Hayashi and colleagues

(Hayashi et al., 2000) observed activation of right frontal areas during subtraction but not multiplication, which they attributed to greater working memory demands in their subtraction task. In the current study, the task design was the same for the transposing and calculation tasks; nevertheless, differences in visual display and the nature of mental conversion needed to solve each item may have placed more demands on working memory in transposing compared to math calculations.

Along with left and right medial frontal areas, several other frontal lobe regions were highly active in the current experimental tasks. The right inferior frontal lobe has been described as important for supporting mathematical subtraction (Kong et al., 2005). These results, along with high activity of the right inferior frontal area in the 5T task, occipital connections to right inferior frontal cortex in the 1T task (shown in neural synchrony analysis), and high activity in left inferior frontal area in the 1D task suggest that these inferior frontal areas should be examined further in comparisons of neural activity in music versus math. The specific roles of the right superior frontal gyrus (highly active in both 1D and 1T), and the right medial orbitofrontal area (highly active in 1T) in working memory also need further study as they relate to musical and mathematical cognition. Given that activation patterns of frontal lobe and frontal-posterior networks can be influenced by specific modifications to task design within domain (e.g., differential effects of a pre-cue versus a post-cue; Ruff et al., 2007), more research is needed to define how task adjustments within and across music and math tasks may affect similarities and differences in corresponding neural activity.

Neural responses elicited during the calculation tasks in the current study are affected by the high working memory demands of the tasks. Using an arithmetic equation verification task in which working memory demands were limited, Rosenberg-Lee and colleagues (Rosenberg-Lee et al., 2011) assessed neural responses in tasks of addition, subtraction, multiplication and division. Among the brain regions they identified as involved in these tasks, the posterior parietal cortex (PPC) was shown to be critically involved in representing cognitive processes of retrieval, calculation and inversion differentially involved in these tasks. Others have identified the PPC as important in supporting mathematical functions (Dehaene and Cohen, 1997). In the current study, MEG was sensitive to differences in high and low working memory load across tasks, as described above. However, all of the tasks were demanding of working memory and did result in high activity in brain regions supporting working memory and, overall, relatively lower activation of brain regions supporting other cognitive elements of the tasks.

In the present investigation, though participants were required to silently name the solutions to each transposing or math problem (e.g., “B” or “7” for the 1T and 1D tasks, and sequentially during the 5T and 5D tasks), left hemisphere activation was not dominant during these tasks. This is in contrast to the results of laterality index analyses in our previous study (Lu et al., 2019) in which silent naming of English letters and silent reading of notes in musical sight-reading both resulted in left hemisphere-dominant activation. We (Lu et al., 2019) hypothesized that those results reflected left hemisphere phonological activation

to support silent naming based on prior findings that MEG neural activity supporting silent naming is lateralized to the left hemisphere in right-handed individuals (Bowyer et al., 2004). In the current study, results from the three analysis methods used do not show that brain activation is more lateralized to the left hemisphere for the transposing or calculation tasks. This is another example of how high task demands on working memory here may have eclipsed other cognitive elements of the tasks that can be observed using MEG when the working memory task demands are lower.

Expert musicians, and students being trained in musical transposing, may employ one or more strategies to accomplish transposing during ongoing performance. We previously described potential differences in musical transposing strategies that may include reliance on auditory imagery, visual-spatial imagery, or both (Lu et al., 2021). The activation of motor cortex during the 5T task in the current study may reflect a motor strategy in transposing. It remains unclear whether this motor activation reflects motor intention or occurred unconsciously in these highly trained musicians presented with musical notation. Further research is warranted into individual transposing strategies used by musicians, and by students during their development of transposing skills. The possibility of strategic motor involvement in transposing could be examined for potential differences in motor activation in response to notes presented in the treble clef (associated with right hand movement in piano) versus bass clef (associated with left hand movement in piano).

Current research into whether music training can affect math achievement includes hypotheses about the effects of music training on domain-general processes such as working memory (Eriksson et al., 2015). These effects have been examined in relation to subtypes of working memory; for example, Roden and colleagues (Roden et al., 2012) found that music training resulted in no improvement in visual memory, but that verbal memory improved. Simmons and colleagues (Simmons et al., 2012) described that subcomponents of working memory have different relationships with different mathematical skills. Continued research into cognitive elements involved in easier and harder iterations of music and math tasks (and cognitive elements shared across tasks) will influence efforts to specify the neural correlates of these cognitive subcomponents. MEG tasks with lower working memory load may allow observation of neural activity supporting specific cognitive elements that are not seen when working memory load is high; however, an advantage of using tasks with high memory load is that they may be closer to conditions in the real world. In real world performance of music or math, working memory demands will vary with the task but will be generally high. For example, during real world musical performance, the broader influences of context and attentional demands along with the musician's work to interpret meaning and emotion all contribute to task complexity. With increasing expertise in music or math, information processing load will be reduced in some aspects, such as greater ability to use context to anticipate continuations in music (Waters et al., 1997). Along with ongoing advances in technology and data analysis methods, systematic adjustments of

cognitive task demands on working memory and other cognitive subcomponents will be needed to capture how the brain supports music and math functions in experts and non-experts, and the cognitive effects of training.

CONCLUSION

MEG was sensitive to differences in working memory load during musical transposing and calculation tasks in a group of classically trained musicians. Frontal-occipital connections were highly active during the transposing tasks, but not during the math calculation tasks. Right temporal regions were highly active in the more difficult condition of the calculation task. Multiple frontal lobe regions were highly active across tasks; notably, the left medial frontal area was highly active in all four tasks, but the right medial frontal area was highly active only during calculations. Right motor and premotor regions were highly active in the more difficult condition of the transposing task but not during calculations. Coherence analyses and neural synchrony analyses yielded several similarities in brain activation across the calculation tasks, but latency analyses were sensitive to differences in task complexity across the two tasks due to the high temporal resolution of MEG. As was done in the current study, future studies should compare brain activity in cognitive tasks involving higher versus lower working memory load. Systematic manipulations to task demands on working memory and other cognitive elements of music and math tasks will be necessary to specify the brain regions supporting elements of these tasks in experts and non-experts. MEG is sensitive to these effects and can be used to examine the neural correlates of musical and mathematical cognition and the consequences of music training.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available as the format is unique. Requests to access the datasets should be directed to C-IL, chingilu@gmail.com.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Taipei Veterans General Hospital Human Research Review Board, Taipei. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

C-IL, MG, and SB contributed to conception and design of the study. C-IL and Y-YL contributed to study procedures, and collection and organization of data. C-IL and SB processed the data. C-IL performed the statistical analyses. C-IL and MG wrote sections of the manuscript. MG wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

REFERENCES

- Aminoff, E. M., Kveraga, K., and Bar, M. (2013). The role of the parahippocampal cortex in cognition. *Trends Cogn. Sci.* 17, 379–390. doi: 10.1016/j.tics.2013.06.009
- Barton, B., and Brewer, A. A. (2013). Visual working memory in human cortex. *Psychology(Irvine)* 4, 655–662. doi: 10.4236/psych.2013.48093
- Bergee, M., and Weingarten, K. M. (2021). Multilevel models of the relationship between music achievement and reading and math achievement. *J. Res. Music Educ.* 68, 398–418. doi: 10.1177/0022429420941432
- Bowyer, S. M. (2016). Coherence a measure of the brain networks: past and present. *Neuropsychiatr. Electrophysiol.* 2:1. doi: 10.1186/s40810-015-0015-7
- Bowyer, S. M., Moran, J. E., Mason, K. M., Constantinou, J. E., Smith, B. J., Barkley, G. L., et al. (2004). MEG localization of language-specific cortex utilizing MR-FOCUSS. *Neurology* 62, 2247–2255. PMIsD:NOPMID doi: 10.1212/01.wnl.0000130385.21160.7a
- Dehaene, S., and Cohen, L. (1997). Cerebral pathways for calculation: double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex* 33, 219–250. doi: 10.1016/s0010-9452(08)70002-9
- Duncan, J., and Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci.* 23, 475–483. doi: 10.1016/S0166-2236(00)01633-7
- Elisevich, K., Shukla, N., Moran, J. E., Smith, B., Schultz, L., Mason, K., et al. (2011). An assessment of MEG coherence imaging in the study of temporal lobe epilepsy. *Epilepsia* 52, 1110–1119. doi: 10.1111/j.1528-1167.2011.02990.x
- Eriksson, J., Vogel, E. K., Lansner, A., Bergstrom, F., and Nyberg, L. (2015). Neurocognitive architecture of working memory. *Neuron* 88, 33–46. doi: 10.1016/j.neuron.2015.09.020
- Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). Mini-Mental state: a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198. doi: 10.1016/0022-3956(75)90026-6
- George, E. M., and Coch, D. (2011). Music training and working memory: an ERP study. *Neuropsychologia* 49, 1083–1094. doi: 10.1016/j.neuropsychologia.2011.02.001
- Goodale, M. A., and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25. doi: 10.1016/0166-2236(92)90344-8
- Hayashi, N., Ishii, K., Kitagaki, H., and Kazui, H. (2000). Regional differences in cerebral blood flow during recitation of the multiplication table and actual calculation: a positron emission tomography study. *J. Neurol. Sci.* 176, 102–108. doi: 10.1016/S0022-510X(00)00323-3
- Hille, A., and Schupp, J. (2015). How learning a musical instrument affects the development of skills. *Econ. Educ. Rev.* 44, 56–82. doi: 10.1016/j.econedurev.2014.10.007
- Holochwost, S. J., Propper, C. B., Wolf, D. P., Willoughby, M. T., Fisher, K. R., Kolacz, J., et al. (2017). Music education, academic achievement, and executive functions. *Psychol. Aesthet. Creat. Arts* 11, 147–166. doi: 10.1037/aca0000112
- Ivanova, A. A., Srikant, S., Sueoka, Y., Kean, H. H., Dhamala, R., O'Reilly, U., et al. (2020). Comprehension of computer code relies primarily on domain-general executive brain regions. *Elife* 9:e58906. doi: 10.7554/eLife.58906
- Jeneson, A., and Squire, L. (2012). Working memory, long-term memory, and medial temporal lobe function. *Learn. Mem.* 19, 15–25. doi: 10.1101/lm.024018.111
- Kong, J., Wang, C., Kwong, K., Vangel, M., Chua, E., and Gollub, R. (2005). The neural substrate of arithmetic operations and procedure complexity. *Cogn. Brain Res.* 22, 397–405. doi: 10.1016/j.cogbrainres.2004.09.011
- Lajiness-O'Neill, R., Richard, A. E., Moran, J. E., Olszewski, A., Pawluk, L., Jacobson, D., et al. (2014). Neural synchrony examined with magnetoencephalography (MEG) during eye gaze processing in autism spectrum disorders: preliminary findings. *J. Neurodev. Disord.* 6:15. doi: 10.1186/1866-1955-6-15
- Lu, C., Greenwald, M. L., Lin, Y., and Bowyer, S. M. (2019). Reading musical notation versus english letters: mapping brain activation with MEG. *Psychol. Music* 47, 255–269. doi: 10.1177/0305735617744886
- Lu, C., Greenwald, M. L., Lin, Y., and Bowyer, S. M. (2021). Musical transposing versus sight-reading: mapping brain activation with Magnetoencephalography (MEG). *Psychol. Music* 49, 581–599. doi: 10.1177/0305735619883692
- Mishkin, M., and Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behav. Brain Res.* 6, 57–77. doi: 10.1016/0166-4328(82)90081-X
- Mongelli, V., Dehaene, S., Vinckier, F., Peretz, I., Bartolomeo, P., and Cohen, L. (2017). Music and words in the visual cortex: the impact of musical expertise. *Cortex* 86, 260–274. doi: 10.1016/j.cortex.2016.05.016
- Moran, J. E., Bowyer, S. M., and Tepley, N. (2005). Multi-resolution FOCUSS: a source imaging technique applied to MEG data. *Brain Topogr.* 18, 1–17. doi: 10.1007/s10548-005-7896-x
- Nakai, T., and Okanoya, K. (2018). Neural evidence of cross-domain structural interaction between language and arithmetic. *Sci. Rep.* 8:12873. doi: 10.1038/s41598-018-31279-8
- Nutley, S. B., Darki, F., and Klingberg, T. (2014). Music practice is associated with development of working memory during childhood and adolescence. *Front. Hum. Neurosci.* 7:926. doi: 10.3389/fnhum.2013.00926
- 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
- Ren, Z., Zhang, Y., He, H., Feng, Q., Bi, T., and Qui, J. (2019). The different brain mechanisms of object and spatial working memory: voxel-based morphometry and resting-state functional connectivity. *Front. Hum. Neurosci.* 13:248. doi: 10.3389/fnhum.2019.00248
- Roden, L., Kreutz, G., and Bongard, S. (2012). Effects of a school-based instrumental music program on verbal and visual memory in primary school children: a longitudinal study. *Front. Psychol.* 3:572. doi: 10.3389/fpsyg.2012.00572
- Rosenberg-Lee, M., Chang, T. T., Young, C. B., Wu, S., and Menon, V. (2011). Functional dissociations between four basic arithmetic operations in the human posterior parietal cortex: a cytoarchitectonic mapping study. *Neuropsychologia* 49, 2592–2608. doi: 10.1016/j.neuropsychologia.2011.04.035
- Ruff, C. C., Blankenburg, F., Bjoertom, O., Bestmann, S., Freeman, E., Haynes, J. D., et al. (2006). Concurrent TMS-fMRI and psychophysics reveal frontal influences on human retinotopic visual cortex. *Curr. Biol.* 16, 1479–1488. doi: 10.1016/j.cub.2006.06.057
- Ruff, C. C., Kristjansson, A., and Driver, J. (2007). Readout from iconic memory and selective spatial attention involve similar neural processes. *Psychol. Sci.* 18, 901–909. doi: 10.1111/j.1467-9280.2007.01998.x
- Sala, G., and Gobet, F. (2020). Cognitive and academic benefits of music training with children: a multilevel meta-analysis. *Mem. Cognit.* 48, 1429–1441. doi: 10.3758/s13421-020-01060-2
- Schön, D., Anton, J. L., Roth, M., and Besson, M. (2002). An fMRI study of music sight-reading. *Neuro Rep.* 13, 2285–2289. doi: 10.1097/01.wnr.0000044224.79663.f5
- Simmons, F., Willis, C., and Adams, A. (2012). Different components of working memory have different relationships with different mathematical skills. *J. Exp. Child Psychol.* 111, 139–155. doi: 10.1016/j.jecp.2011.08.011
- Sligte, L. G., Scholte, H. S., and Lamme, V. A. F. (2009). V4 activity predicts the strength of visual short-term memory representations. *J. Neurosci.* 29, 7432–7438. doi: 10.1523/JNEUROSCI.0784-09.2009
- Stewart, L. (2005). A neurocognitive approach to music reading. *Ann. N. Y. Acad. Sci.* 1060, 377–386.
- Stewart, L., Henson, R., Kampe, K., Walsh, V., Turner, R., and Frith, U. (2003). Brain changes after learning to read and play music. *Neuroimage* 20, 71–83. doi: 10.1016/S1053-8119(03)00248-9
- Swaminathan, S., Schellenberg, E., and Khalil, S. (2017). Revisiting the association between music lessons and intelligence: training effects or music aptitude? *Intelligence* 62, 119–124. doi: 10.1016/j.intell.2017.03.005
- Van de Cavey, J., and Hartsuiker, R. (2016). Is there a domain-general cognitive sequencing system? Evidence from structural priming across music, math, action descriptions, and language. *Cognition* 146, 172–184. doi: 10.1016/j.cognition.2015.09.013
- Wallick, M. (1998). A comparison of ohio proficiency test results between fourth-grade pullout students and those of matched ability. *J. Res. Music Educ.* 46, 239–247. doi: 10.2307/3345626
- Waters, A. J., Underwood, G., and Findlay, J. M. (1997). Studying expertise in music reading: use of a pattern-matching paradigm. *Percept. Psychophys.* 59, 477–488. doi: 10.3758/BF03211857

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale*, 3rd Edn. San Antonio, TX: The Psychological Corp.

Weiner, K. S., Barnett, M. A., Lorenz, S., Caspers, J., Stigliani, A., Amunts, K., et al. (2017). The cytoarchitecture of domain-specific regions in human high-level visual cortex. *Cereb. Cortex* 27, 146–161. doi: 10.1093/cercor/bhw361

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 reviewer EWP declared a past co-authorship with one of the authors SB to the handling editor.

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.

Copyright © 2022 Lu, Greenwald, Lin and Bowyer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Usability of a Hybrid System Combining P300-Based Brain-Computer Interface and Commercial Assistive Technologies to Enhance Communication in People With Multiple Sclerosis

Angela Riccio^{1,2*†}, Francesca Schettini^{1,2†}, Valentina Galiotta¹, Enrico Giraldi¹, Maria Grazia Grasso³, Febo Cincotti⁴ and Donatella Mattia^{1,2}

OPEN ACCESS

Edited by:

Davide Valeriani,
Neurable Inc., United States

Reviewed by:

Sergei L. Shishkin,
Moscow State University
of Psychology and Education, Russia
Caterina Cinel,
University of Essex, United Kingdom

*Correspondence:

Angela Riccio
a.riccio@hsantalucia.it

[†] These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Brain-Computer Interfaces,
a section of the journal
Frontiers in Human Neuroscience

Received: 02 February 2022

Accepted: 26 April 2022

Published: 26 May 2022

Citation:

Riccio A, Schettini F, Galiotta V,
Giraldi E, Grasso MG, Cincotti F and
Mattia D (2022) Usability of a Hybrid
System Combining P300-Based
Brain-Computer Interface
and Commercial Assistive
Technologies to Enhance
Communication in People With
Multiple Sclerosis.
Front. Hum. Neurosci. 16:868419.
doi: 10.3389/fnhum.2022.868419

¹ Neuroelectric Imaging and BCI Lab, Fondazione Santa Lucia (IRCCS), Rome, Italy, ² Servizio Ausilioteca per la Riabilitazione Assistita con Tecnologia, Fondazione Santa Lucia (IRCCS), Rome, Italy, ³ Multiple Sclerosis Unit, Fondazione Santa Lucia (IRCCS), Rome, Italy, ⁴ Department of Computer, Control and Management Engineering Antonio Ruberti, Sapienza University of Rome, Rome, Italy

Brain-computer interface (BCI) can provide people with motor disabilities with an alternative channel to access assistive technology (AT) software for communication and environmental interaction. Multiple sclerosis (MS) is a chronic disease of the central nervous system that mostly starts in young adulthood and often leads to a long-term disability, possibly exacerbated by the presence of fatigue. Patients with MS have been rarely considered as potential BCI end-users. In this pilot study, we evaluated the usability of a hybrid BCI (h-BCI) system that enables both a P300-based BCI and conventional input devices (i.e., muscular dependent) to access mainstream applications through the widely used AT software for communication “Grid 3.” The evaluation was performed according to the principles of the user-centered design (UCD) with the aim of providing patients with MS with an alternative control channel (i.e., BCI), potentially less sensitive to fatigue. A total of 13 patients with MS were enrolled. In session I, participants were presented with a widely validated P300-based BCI (P3-speller); in session II, they had to operate Grid 3 to access three mainstream applications with (1) an AT conventional input device and (2) the h-BCI. Eight patients completed the protocol. Five out of eight patients with MS were successfully able to access the Grid 3 via the BCI, with a mean online accuracy of 83.3% (± 14.6). Effectiveness (online accuracy), satisfaction, and workload were comparable between the conventional AT inputs and the BCI channel in controlling the Grid 3. As expected, the efficiency (time for correct selection) resulted to be significantly lower for the BCI with respect to the AT conventional channels ($Z = 0.2$, $p < 0.05$). Although cautious due to the limited sample size, these preliminary findings indicated that the BCI control channel did not have a detrimental effect with respect to conventional AT channels on the ability to operate an

AT software (Grid 3). Therefore, we inferred that the usability of the two access modalities was comparable. The integration of BCI with commercial AT input devices to access a widely used AT software represents an important step toward the introduction of BCIs into the AT centers' daily practice.

Keywords: assistive technologies, brain-computer interface, multiple sclerosis, P300, Grid 3, end-users, user-centered design, usability

INTRODUCTION

Multiple sclerosis (MS) is an autoimmune disease characterized by clinical neurological relapses and progressive loss of motor and sensory function that affects approximately 2.8 million people worldwide (MSIF, 2020). The course of MS is highly variable, but the relapsing and/or progressive course of the disease leads to a long-term sensorimotor disability (Oh et al., 2018). The level of disability can be even magnified by the presence of a characteristic symptom occurring in MS, such as fatigue (Tur, 2016). Fatigue is indeed one of the most common symptoms and is present in almost 80% of the patients with MS (Rottoli et al., 2017), and it can be severe in up to 60% of patients (Hadjimichael et al., 2008). The impact on quality of life (QoL) of such MS long-term consequences is considerably high, especially if one considers the relatively young age of the population affected by MS (Compston and Coles, 2008). Motor disability and fatigue in MS may result in substantial impairment in communication and in the access to digital technologies, thus leading to overall social isolation.

Assistive technology (AT) indicates any product that enables people of all ages with activity limitations in their daily life, education, work, or leisure (Andrich et al., 2013). ATs include various input devices (e.g., mouse emulators, eye-trackers, adapted joysticks, and speech recognition) and specific software (e.g., Grid 3, Smartbox Assistive Technology, 2021) to create customized solutions to overcome disability. ATs are selected and customized based on users' needs and their motor, sensory, and cognitive impairment (disabilities) and are validated according to the user-centered design (UCD; ISO, 2019) that is defined as an iterative process that involves end-users in all the stages of technology design, development, and testing.

ATs in general can support communication and environmental interaction in people with disabilities due to MS; however, since all conventional AT input devices are muscular dependent, their usability may result to be compromised by the presence of muscular fatigue in these patients (Tur, 2016).

The brain-computer interface (BCI) technology has been demonstrated to provide severely (motor) disabled people with an alternative channel to enhance/restore communication and environmental control that is independent from the physiological peripheral pathways (i.e., nerves and muscles) (Nijboer et al., 2008; Sellers et al., 2010; McCane et al., 2014; Riccio et al., 2015; Schettini et al., 2015; Guy et al., 2018; Wolpaw et al., 2018; Medina-Juliá et al., 2020). Most of the current studies on the feasibility and usability of non-invasive BCIs systems for communication have relied on evoked potentials (EPs) and event-related potentials (ERPs) (e.g., N200; Treder and Blankertz, 2010) as control features (Powers et al., 2015; Allison et al., 2020).

More recently, the so-called hybrid BCIs (h-BCIs) have been proposed that utilize more than one physiological signal and/or external signals to increase, for instance, the accuracy and/or the information transfer rate (Choi et al., 2017). The role of these h-BCIs appears particularly relevant in the domain of AT as they can be conceived as an additional input to provide multimodal access (BCI and conventional AT input devices) to AT software for communication and environmental control functionalities (Millán et al., 2010; Riccio et al., 2011, 2015, 2016; Zickler et al., 2011; Thompson et al., 2014; Schettini et al., 2015). The incorporation of the BCI as an input channel to commercial AT software becomes also essential to improve BCI modularity (Liberati et al., 2015) and eventually to better adapt it to users' sensory, cognitive, and motor profiles (Schreuder et al., 2013).

Till present, patients with MS have been rarely considered as potential end-users of BCIs to support communication/interaction. Martinez-Cagigal et al. (2016) evaluated a P300-based BCI to access a web browser to eventually support communication in patients with MS.

In this study, we evaluated the usability of a newly implemented AT system that adds a P300-based BCI technology to commercial AT input devices (h-BCI system) to eventually access a range of computer applications through a commercial comprehensive AT software for communication and interaction, the Grid 3 platform (Smartbox Assistive Technology, 2021), with the aim to exploit the BCI as an additional input to ATs to address the issues of fatigue limiting the everyday use of ATs in MS end-users with different degrees of disability. The h-BCI system was evaluated according to the UCD metrics and, therefore, in terms of effectiveness, efficiency, and satisfaction.

As such, the proposed hybrid combination of a P300-based BCI with different conventional input devices can eventually enable patients with MS to switch between a muscular-based channel (e.g., joystick control, mouse control, and head tracker control) and a P300-based BCI channel, according to the level of fatigue they experience or their preference. The integration of this h-BCI system with the Grid 3 platform can guarantee universal access to every kind of mainstream application running on a PC (i.e., browsing the Internet and WhatsApp).

MATERIALS AND METHODS

Participants and Routine Clinical Assessment

A total of 13 participants with MS, according to revised McDonald criteria (Thompson et al., 2018), were enrolled in

the study (patients with MS; mean age \pm SD = 51.6 \pm 12.9; two women; mean time since diagnosis: 253.4 months, range: 70–399 months).

The inclusion criteria were (1) ≥ 18 years, (2) diagnosis of MS, and (3) functional limitation in at least one aspect of interpersonal communication or environmental interaction.

The exclusion criteria were (1) global cognitive decline, (2) concomitant aphasia or comprehension deficits, (3) visual field deficits, (4) severe concomitant medical conditions (e.g., fever and infections), and (5) periods of disease exacerbation.

All participants (or their legal guardians when necessary) gave their written informed consent for participation in the study. The study was approved by the Local Ethical Committee (CE/PROG.707) of Fondazione Santa Lucia, IRCCS.

Patients with MS were recruited from those admitted to the AT service of Fondazione Santa Lucia, IRCCS (Rome, Italy), because of their limitations in at least one aspect related to communication and/or environmental interaction. According to the clinical standard care, they underwent a neuropsychological assessment (see **Table 1**) and were administered with the Fatigue Severity Scale (FSS; Krupp et al., 1989) to assess the severity of fatigue symptoms (scores range: 1–7; low fatigue-high fatigue; mean \pm SD = 47.1 \pm 10.6) and the Expanded Disability Status Scale (EDSS; Kurtzke, 1983) to quantify the level of physical and cognitive disability in MS (scores range: 1–10; normal neurological exam-death; mean \pm SD = 7.1 \pm 2.8). When enrolled in the study, all patients were already using an AT device/solution based on their needs [see Individually Prioritized Problem Assessment (IPPA) and AT in **Table 1**]. All patients with MS were naïve to BCI protocol.

Hybrid Brain-Computer Interface System

We implemented an h-BCI system prototype based on the communication between the AT software Grid 3 (Smartbox Assistive Technology, 2021) and the BCI software BCI2000 (Schalk et al., 2004). The system combines the P300 ERP with conventional input devices (e.g., head tracker and mouse) as input channels, thus resulting in a hybrid control of the system (Grid 3).

Grid 3 is a highly versatile AT software used to create customized interfaces, providing aided access to PC applications, and allowing to combine different input devices (e.g., head tracker and switch). It is one of the most used commercial AT software for communication in AT services. Grid 3 accepts inputs from conventional keyboards and allows to associate a maximum of eight keys to specific and customizable actions to control applications (e.g., F1 can act as “back to home page,” F2 as “jump to keyboard grid,” and F3 as “turn up the volume”). Taking advantage of this feature, we modified the BCI2000 source code (Schalk et al., 2004) so that selections made with the P300-speller application generated a keypress event and activated specific commands associated with the key by means of Grid 3. By doing so, Grid 3 operated as a “link” between BCI2000 and the specific applications (**Figure 1**; e.g., WhatsApp, YouTube, and Google Chrome).

To carry out the experimental protocol, we used a 34-inch screen (3,440 \times 1,440 pixel, 79.7 \times 33.3 cm). The distance

between the screen and the patient’s eyes was 100 cm. Grid 3 interfaces were characterized by the presence of a control matrix on the left side of the screen and the application interface on the right side of the screen, where commands were delivered. The control matrix for each application consisted of the main starting matrix (**Figures 2A–C**) and of a keyboard matrix (6 \times 5; **Figure 2D**), which could be accessed by selecting a specific icon on the main matrix (**Figure 2**). The interface delivered auditory feedback when a selection was made on control matrices.

WhatsApp control matrix (3 \times 3; **Figure 2A**) allowed to move through contacts and messages and to write and send messages, emojis, and predefined messages. Emojis and predefined messages were organized in two different submatrices (4 \times 3); in the h-BCI system control, the latter matrices were accessible by means of a conventional input device (hybrid control).

YouTube control matrix (4 \times 3; **Figure 2B**) allowed to search videos, go through a list of videos, play a video, and control the volume.

Google Chrome control matrix (4 \times 3; **Figure 2C**) allowed to make personalized surfing on the web, scan and select links on a web page, go back to a previous page, and refresh a web page. Each matrix included a command to pause the BCI control.

Evaluation Protocol

The protocol consisted of two sessions performed on 2 different days.

Session I – BCI control ability test: Patients with MS were asked to control a P300-speller (Farwell and Donchin, 1988) to familiarize themselves with a BCI system and test their ability to control a validated BCI system whose reliability has been largely demonstrated (Rezeika et al., 2018). The P300-speller session lasted about 90 min.

Session II – Grid 3 access via AT input and via BCI control: Participants were asked to use Grid 3 to access three applications (WhatsApp, YouTube, and Web Browser) with (1) a conventional commercial AT input device (Grid 3 access via a conventional AT input) and (2) with the h-BCI system (Grid 3 access via BCI control). Interfaces and tasks were comparable in the two conditions. Data obtained in the two conditions were compared in terms of usability. This session lasted about 120 min.

Sessions I and II were performed within 10 days.

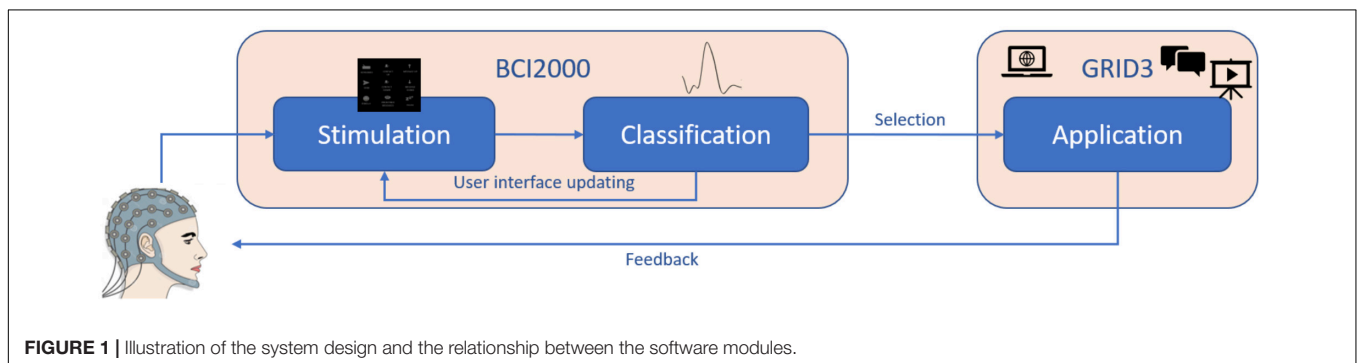
Session I: Brain-Computer Interface Control Ability Test

Patients with MS had to operate a P300-speller (Farwell and Donchin, 1988) on a 15” screen. The distance between the screen and the patient’s eyes was 60 cm. Scalp EEG signals were acquired using a 16-channel amplifier (g.USBamp, g.tec, Austria; 256 Hz) from 16 sintered Ag/AgCl active electrodes (g.Ladybird, g.tec, Austria) placed according to the 10-10 International System (Fz, Cz, Pz, Oz, P3, P4, PO7, PO8, F3, F4, FCz, C3, C4, CP3, CPz, CP4, referenced to the right ear lobe and grounded to the left mastoid). A conductive gel was applied between the electrodes and the scalp to lower impedances. The impedance value did not exceed 10 k Ω . The P300-speller consisted of a 6 \times 5 matrix, which contained 30 alphabetic characters intensified in pseudo-random

TABLE 1 | Information about the neuropsychological assessment, the problems, the assistive technology, and the participation in the h-BCI evaluation for each patient included in the study.

MS patient	Neuropsychological assessment	Problems (IPPA)	Assistive Technology	h-BCI
P1	Executive Functions: ✓ Attention: ✓ Working Memory: ✓	Computer accessibility	Mouse emulator: Joystick	✓
P2	Executive Functions: ✗ Attention: ✓ Working Memory: ✗	Reading due to fatigue	Mainstream solutions with customized accessibility settings	✓
P3	Executive Functions: ✓ Attention: ✓ Working Memory: ✓	Computer and smartphone accessibility	Mouse emulator: Head-tracker	✓
P4	Executive Functions: ✓ Attention: ✓ Working Memory: ✓	Computer accessibility	Mainstream solutions with customized accessibility settings	✓
P5	-	Computer accessibility	Mainstream solutions with customized accessibility settings	✓
P6	Executive Functions: ✓ Attention: ✓ Working Memory: ✓	-	Mainstream solutions with customized accessibility settings	✓
P7	Executive Functions: ✗ Attention: ✗ Working Memory: ✗	Computer accessibility	Mainstream solutions with customized accessibility settings	✓
P8	Executive Functions: ✗ Attention: ✗ Working Memory: ✗	Face to face communication Reading/writing Making phone calls	Customized Grid 3 interface operated with a button switch and a scanning modality	✓
P9	Executive Functions: ✗ Attention: ✗ Working Memory: ✗	Smartphone accessibility Reading	Mainstream solutions with customized accessibility settings	✗
P10	Executive Functions: ✓ Attention: ✓ Working Memory: ✓	Face to face communication Writing/reading Computer and smartphone accessibility Social interactions Independence in daily activity	Mainstream solutions with customized accessibility settings	✗
P11	Executive Functions: ✓ Attention: ✗ Working Memory: ✓	Computer accessibility Writing/reading	Mouse emulator: Joystick	✗
P12	Executive Functions: ✓ Attention: ✓ Working Memory: ✗	Face to face communication Smartphone accessibility Accessibility to entertainment applications.	Customized Grid 3 interface to support access to PC applications, operated with a button switch and a scanning modality	✗
P13	-	Computer and smartphone accessibility Accessibility to domotic system	Head tracker to support access to PC	✗

Neuropsychological assessment: the column reports the results of clinical neuropsychological assessment for executive functions, attention, and working memory. The mark "✗" stands for a deficit, the mark "✓" stands for normal cognitive functioning, and "-" stands for the absence of a neuropsychological assessment. Problems (IPPA): the column reports the problems identified by participants by filling in the IPPA (Individually Prioritized Problem Assessment interview; Wessels et al., 2002). IPPA is a semi-structured interview that aims at investigating seven (or fewer) problems (related to communication and environmental interaction in this case) that the patients would like to solve with the AT. Assistive Technology: the column reports the AT solution used by each participant to overcome the problems identified with the IPPA. H-BCI: the column reports the participation (or not) in the h-BCI system evaluation session. The mark "✗" is used when the patient completed all the sessions included in the protocol, and the mark "✓" is used when the patients did not complete all the tasks included in the protocol.



groups according to the checkerboard stimulation paradigm (Townsend et al., 2010). The stimulus duration was 125 ms and the interstimulus interval (ISI) was 125 ms [stimulus onset interval (SOA): 250 ms].

Participants had to spell eight predefined five-character words (eight runs) with a pause of about 3 min between them. Each target was intensified 16 times, corresponding to eight stimulation sequences, and participants were instructed to attend the target stimulus and mentally count how many times it was intensified. Target characters were cued at the beginning of each trial. No feedback was provided during the first three runs (15 selections; calibration). This EEG data set was used to extract the BCI classifier parameters by applying a Stepwise Linear Discriminant Analysis (SWLDA; Krusienski et al., 2006). We used the BCI classifier parameters to determine the online feedback on the spelling of the following five words (25 selections; online copy mode).

Session II: Grid 3 Access *via* Assistive Technology Input and *via* Brain-Computer Interface Control

Grid 3 Access via Assistive Technology Input

The “Grid 3 access *via* AT input” condition was aimed at evaluating patients’ ability to operate Grid 3 independently from the BCI channel. Patients with MS operated Grid 3 with their own conventional (PC mouse) or alternative input device (e.g., head tracker and switch; see **Table 1**).

In this condition, the number of selections needed to complete a task varied as a function of the input channel: e.g., the selection of an item operated with a switch and the scanning modality required a double-action (selection of the row and then of the column containing it) in comparison with the selection performed with the mouse emulators, allowing direct control of the cursor (e.g., head tracker and joystick).

Patients with MS were first instructed about the experimental procedures. They were then asked to perform the following tasks:

- Task 1: WhatsApp (selections required: 8 or 16 in case of scanning modality). Participants had to select the second contact in the chat list by scrolling it down, open the keyboard, and write the message “OK,” then they had to go back to the “main menu,” send the message, open the emoji menu, and send the second emoji.
- Task 2: YouTube (selections required: 5 or 10 in case of scanning modality). Participants were required to scroll down the video list and open the second video, turn up the volume, and pause the video.
- Task 3: Google Chrome (selections required: 5 or 10 in case of scanning modality). Participants had to open the second link of a web page, scroll down the page, and select the command “pause.”

Grid 3 Access via Brain-Computer Interface Control

Patients with MS had to operate Grid 3 using the h-BCI, combining the P300-based BCI and the conventional/alternative input device used in the Grid 3 access *via* AT input condition (hybrid control). Scalp EEG signals were acquired using a 16-channel amplifier (g.USBamp, g.tec, Austria; 256 Hz) from eight

sintered Ag/AgCl active electrodes (g.Ladybird, g.tec, Austria) placed according to the 10-10 International System (Fz, Cz, Pz, Oz, P3, P4, PO7, PO8, referenced to the right ear lobe and grounded to the left mastoid). A conductive gel was applied between the electrodes and the scalp to lower impedances. Each h-BCI system session consisted of a “Calibration” and an “Online mode.” Patients with MS were instructed to focus their attention on the target and mentally count how many times it was intensified. The experimenter pointed at the target stimuli.

Calibration consisted of six runs (4 items each; 24 total items) with matrices of different sizes: two runs with a 3×3 matrix, two runs with a 4×3 matrix, and two runs with a 6×5 matrix. No feedback was provided to participants. Items were randomly intensified by rows and columns for 125 ms and with an ISI of 125 ms; each item was intensified 30 times (15 stimulation sequences). The EEG data set collected during the calibration was used to extract the BCI classifier parameters by applying an SWLDA.

The “online mode” consisted of three runs; patients were asked to perform the same tasks performed during the “Grid 3 access *via* AT input” condition. The number of stimulation sequences was optimized for each participant. The criterion applied to establish the number of sequences to be used in the online mode was $n + 1$, where “ n ” was the number of sequences necessary to reach 100% offline accuracy (applying a sixfold cross-validation procedure to the data recorded in the calibration). Feedback occurred at the end of each trial.

Task 1: WhatsApp (minimum number of selections required: 8 or 9 in case of hybrid control based on scanning modality). Only the emoji (one out of eight selections) was selected with the hybrid control (conventional/alternative input device); the remaining seven were BCI-based selections. This leads to variability in the number of selections.

Task 2: YouTube (minimum number of selections required: 5).

Task 3: Google Chrome (minimum number of selections required: 5).

In case of wrong selection, the participant had to correct the error.

The tasks were interrupted if the participants reached the $k \times 3$ number of selections (where “ k ” is the minimum number of selections expected to complete the task).

Outcome Measures

Session I: Brain-Computer Interface Control Ability Test

P300-speller performance was evaluated as follows and then compared with h-BCI system performance:

- Online accuracy (%) is defined as the ratio between correct selections and selections needed to complete the task.
- The highest written symbol rate (WSR; Furdea et al., 2009) was assessed as a function of the number of stimulus repetitions delivered in a given trial of the five online copy-mode runs. The maximum WSR value for each subject provides an objective evaluation of the system performance by combining the accuracy level with the time needed to reach it, in terms of the number of stimulation sequences. In

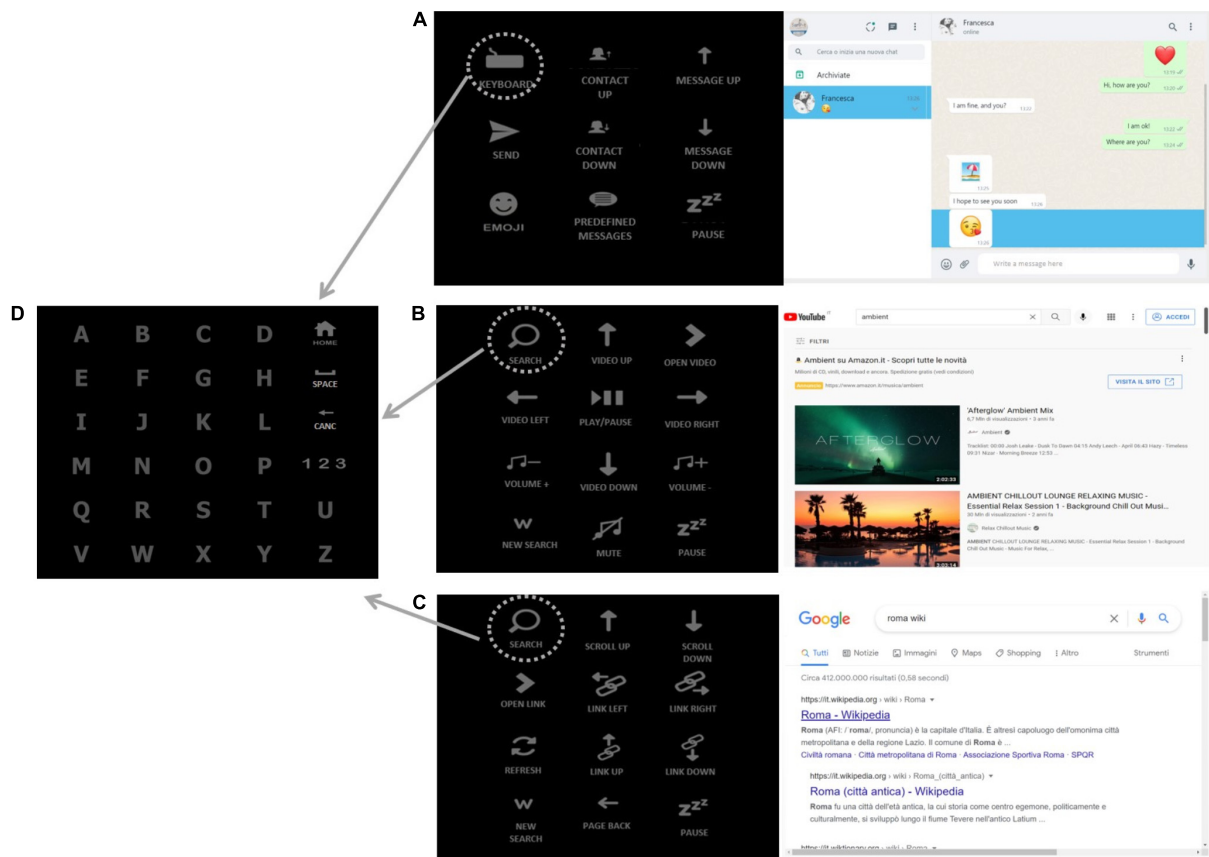


FIGURE 2 | The h-BCI system interfaces. **(A) WhatsApp interface:** “contact up” and “contact down” icons allow the user to scroll through the contacts, the “keyboard icon” (in the upper-left corner of the matrix) allows to open the keyboard **(D)** and spell the message, the message is sent by selecting the “send” icon. “Message up” and “message down” icons allow the user to scroll through the messages received within a contact chat. **(B) YouTube interface:** the “search” icon (in the upper-left corner of the matrix) is selected to search the desired item by opening the keyboard **(D)** and spelling a keyword, the “arrow” icons (video up, video down, video left, and video right) are selected to scroll through the videos displayed on the screen; “open video” icon allows to choose the desired item and the “play/pause” one to start and stop the video. **(C) Google Chrome interface:** the “search” icon (in the upper-left corner of the matrix) is selected to search the desired link, the “scroll up” and “scroll down” icons allow to scroll the links displayed on the screen, “open link” icon allows to choose the link to be opened. **(D) Keyboard:** the keyboard matrix is accessed by selecting the “keyboard” **(A)** and the “search” **(B,C)** icons on the main matrices.

the “copy-mode” runs, the participants were asked to (copy) spell predefined words. In case of errors, the participants were not asked to correct the wrong selection; he/she had to proceed with the next letter of the word. Therefore, the number of selections was fixed.

Session II: Grid 3 Access *via* Assistive Technology Input and *via* Brain-Computer Interface Control

The usability of the system in the two conditions of session II (Grid 3 access *via* AT input and Grid 3 access *via* BCI control) was evaluated in terms of effectiveness, efficiency, and satisfaction (ISO, 2019).

- **Effectiveness** is defined as the accuracy and completeness with which the user achieves goals while using the system. It was evaluated as follows:
- **Online accuracy (%)**: Calculated by dividing the number of correct selections by the total number of selections.

- **Completeness**: Indicated with the number of participants who completed the protocol. The tasks were not considered complete after the $n \times 3$ number of selections (with “n” as the minimum number of selections needed to complete the task).
- **Efficiency** describes the degree to which the system enables quick, effective, and economic performance in terms of time, human effort, costs, and materials. It was evaluated in terms of workload.

The workload was assessed through the National Aeronautics and Space Administration – Task Load Index (NASA-tlx; Hart, 2006). NASA-tlx is a multidimensional questionnaire that assesses perceived workload during the usage of a high technology device. It was administered by an experimenter at the end of both the “Grid 3 access *via* AT input” and “Grid 3 access *via* BCI control” conditions. The overall workload score

TABLE 2 | Performance data in sessions I and II.

MS patient	Session I BCI control ability test (P300-speller)			Session II			
	Acc (%)	WSR	Task completion	Grid 3 access <i>via</i> AT input		Grid 3 access <i>via</i> BCI control	
				Acc (%)	Task completion	Acc (%)	Task completion
P1	56	2.7	✓	100	✓	79.0	✓
P2	88	2.4	✓	87.8	✓	62.8	✓
P3	96	9.1	✓	100	✓	80.4	✓
P4	100	9.4	✓	100	✓	94.4	✓
P5	44	0	✗	100	✓	46.5	✗
P6	92	6	✓	93.3	✓	100	✓
P7	0	0	✗	85.4	✓	11.4	✗
P8	0	0	✗	87.6	✓	40	✗
Mean ± SD	59.5 ± 41.7			94.3 ± 6.5		64.3 ± 30.2	
P9	96	3.4	✓	100%	✓	–	–
P10	96	6.1	✓	–	–	–	–
P13	96	4.9	✓	–	–	–	–
Mean ± SD	69.4						

Acc, accuracy; WSR, written symbol rate. Task completion: the mark “✓” stands for the complete task, the mark “✗” stands for the incomplete task, and the mark “–” means that the patient did not participate in that task.

(0–100) is a weighted average between the rating of six factors (i.e., mental demand, physical demand, temporal demand, performance, effort, and frustration level). Each factor has a weighted rating that ranges from 0 to 33.33. Higher scores are associated with higher levels of workload.

- **Effectiveness/efficiency:** Time per correct selection and WSR were considered metrics belonging to both effectiveness and efficiency constructs:
 - WSR (Furdea et al., 2009) was assessed offline as a function of the number of stimulus repetitions delivered in a given trial of calibration; it was considered the highest offline WSR value. WSR was computed only for the “Grid 3 access *via* BCI control.”
 - Time for correct selection (s) was computed as the ratio between the total time to complete the online tasks and the number of correct selections.
- **Satisfaction** represents the degree to which the user’s physical, cognitive, and emotional responses that result from the use of a system meet the user’s needs and expectations. It was evaluated by means of the System Usability Scale (SUS; Bangor et al., 2009). SUS is a 5-point Likert scale that assesses user satisfaction with a technological device. Participants were required to express their agreement/disagreement with the statements on the scale. It was administered by an experimenter at the end of both conditions.

Statistical Analysis

We investigated the correlation between the P300-speller accuracy and the h-BCI system accuracy by means of Spearman’s rank test. These analyses aimed at demonstrating the reliability

of the h-BCI, assuming the reliability of the well-validated P300-speller.

Usability Assessment

To evaluate whether the introduction of the P300-based BCI as an additional channel to control the AT software would affect system usability, we compared the “Grid 3 access *via* AT input” condition with the “Grid 3 access *via* BCI control” condition for the five patients who successfully controlled the h-BCI. We compared the two conditions in terms of effectiveness (accuracy), efficiency (time per correct selection, NASA-tlx scores), and satisfaction (SUS scores) scores by means of a (non-parametric) Wilcoxon matched-pairs test since the distributions violated the assumption of normality. Regarding the NASA-tlx, we compared both the overall perceived workload score and the single-factor scores (mental demand, physical demand, temporal demand, performance, effort, frustration) by means of a (non-parametric) Wilcoxon matched-pairs test. Finally, any possible correlations between the accuracy in “Grid 3 access *via* BCI control” condition and FSS scores (level of fatigue) and EDSS scores (level of disability) were investigated by means of Spearman’s rank test to evaluate whether the level of fatigue and disability could influence the ability to control the h-BCI.

RESULTS

Pilot Evaluation

We first evaluated the h-BCI system, including a convenience sample of 13 healthy volunteers (mean age = 27.2 ± 2.9 ; nine women), with no history of neurological/psychiatric disorders. All participants were able to control the system with a mean (\pm SD) online accuracy of $98.1 \pm 2.7\%$ and a mean (\pm SD) time per correct selection $25.3 \text{ s} (\pm 8.1)$. The mean (\pm SD) overall

perceived workload (NASA-tlx) was $39.1 (\pm 18.2)$, and the mean (\pm SD) satisfaction score (SUS) was $81.4 (\pm 12.6)$.

Session I: Brain-Computer Interface Control Ability Test

A total of 11 of 13 patients with MS participated in session I (two women; mean age: 52 ± 14.0 ; mean time since diagnosis: 248.8 ± 133.2 months; years of formal education: 13 ± 3.1 years). Three patients gained a control accuracy of the P300-speller below the 50% (Table 2; P5, P7, P8) and showed a WSR = 0, not supporting an efficient communication. Two of those patients had an accuracy of 0%. A total of 8 of 11 patients had a WSR > 2, showing an efficient control of the P300-speller.

Session II: Grid 3 Access via Assistive Technology Input

A total of 9 of 13 patients with MS participated in the “Grid 3 access via AT input” condition (two women; mean age: 54.9 ± 10.6 ; mean \pm SD time since diagnosis: 249.2 ± 128.1 months; mean education: 13 ± 3.5 years).

- **Effectiveness:** All patients with MS completed the tasks (Tables 2, 3), with a mean accuracy of $94.9 \pm 6.4\%$.
- **Efficiency:** Mean time per correct selection of the nine participants in the session was $8.8 \text{ s} (\pm 6.3)$. Total workload (NASA-tlx) was $22.3 (\pm 20.8)$ with weighted ratings of NASA-tlx factors ranging from 1.2 (effort) to 8.1 (mental demand) (Table 4).
- **Satisfaction:** SUS score was on average $78.1 (\pm 10.8)$ (Table 4).

Session II: Grid 3 Access via Brain-Computer Interface Control

Eight patients with MS participated in the “Grid 3 access via BCI control” condition (two women; mean age: 54.1 ± 11.1 ; mean time since diagnosis: 249.5 ± 137.0 months; mean education: 13 ± 3.8 years).

- **Effectiveness.** Five out of eight patients with MS (P1, P2, P3, P4, and P6) controlled the h-BCI system since they completed the three online tasks. Three participants (P5, P7, and P8) did not complete the online tasks (Tables 2, 3). Patients with MS that controlled the h-BCI system obtained a mean online accuracy of $83.3\% \pm 14.6\%$ (Table 3).
- **Efficiency.** The five patients with MS that controlled the system obtained a mean (\pm SD) time per correct selection of $41.0 \text{ s} \pm (16.2)$ and a mean offline WSR of $4.4 \text{ sym/min} (\pm 3.7)$; the mean (\pm SD) overall workload score was $24.4 (\pm 21.0)$ in patients with MS that controlled the h-BCI; weighted ratings of NASA-tlx factors ranged from 1.2 (performance) to 10.7 (mental demand). The mean (\pm SD) overall workload score was $35.8 (\pm 10.9)$ in patients with MS that did not control the h-BCI system, and the weighted ratings of NASA-tlx factors ranged from 1 (physical demand) to 13.3 (mental demand) (Table 4).
- **Satisfaction.** The mean (\pm SD) SUS score (0–100) was $78 (\pm 16.6)$ in the five patients with MS that were able to

control the h-BCI system and $69.2 (\pm 8.0)$ in the three patients with MS who did not control the h-BCI system (Table 4).

Brain-Computer Interface Control Ability Test vs. Grid 3 Access via Brain-Computer Interface Control

Considering all the eight patients with MS who participated in session II, we found a significant correlation between the P300-speller online copy-mode accuracy and the h-BCI system accuracy (Grid 3 access via BCI control condition; $r_s = 0.9$, $p < 0.05$).

The three patients with MS (P5, P7, P8) who did not complete the three tasks in the “Grid 3 access via BCI control” condition also did not control the P300-speller (accuracy < 50%, WSR = 0).

Usability Assessment Effectiveness

All the eight patients who participated in both the “Grid 3 access via AT input” condition and the “Grid 3 access via BCI control” condition successfully completed the “Grid 3 access via AT input” condition. Five of them successfully completed the “Grid 3 access via BCI control” condition (Table 2). Considering the five patients with MS who completed both the conditions, we did not find significant differences in the accuracy ($Z = 1.5$; $p = 0.1$; Figure 3).

No significant correlations were found between the h-BCI system online accuracy and the FSS scores ($r_s = 0.29$, $p = 0.49$) and the EDSS scores ($r_s = -0.57$, $p = 0.013$), respectively.

Efficiency

Considering the five patients with MS who completed both the “Grid 3 access via AT input” condition and the “Grid 3 access via BCI control” condition, we did not find significant differences in workload scores (NASA-tlx scores: overall workload: $Z = 1.1$, $p = 0.3$; mental demand: $Z = 1.1$, $p = 0.3$; physical demand: $Z = 0.9$, $p = 0.4$; temporal demand: $Z = 1.1$, $p = 0.3$; performance: $Z = 0.5$, $p = 0.6$; effort: $Z = 1.8$, $p = 0.1$; frustration: $Z = 1.3$, $p = 0.2$). We found a significant difference in time per correct selection, which was significantly higher in the “Grid 3 access via BCI control” condition ($Z = 2.0$, $p < 0.05$; Figure 3).

Satisfaction

Considering the five participants who successfully controlled the h-BCI system, no significant differences were found between the SUS scores in the “Grid 3 access via AT input” condition and in the “Grid 3 access via BCI control” condition ($Z = 0.9$, $p = 0.3$; Figure 3).

DISCUSSION

We implemented an h-BCI system combining the P300-based BCI technology with commercial AT input devices to access a range of computer applications through a widely used AT software for communication and environmental interaction: Grid 3. We evaluated the usability of the h-BCI system involving

TABLE 3 | The mean (\pm SD) group results, including the patients who completed each session.

Outcome measures		Session I BCI control ability test	Session II	
			Grid 3 access <i>via</i> AT input	Grid 3 access <i>via</i> BCI control
	Patients (n)	8/11	9/9	5/8
Effectiveness	Online accuracy (%)	90 \pm 14.2	94.9 \pm 6.4	83.3 \pm 14.6
Efficiency	NASA-tlx (Total workload; 0–100)	–	22.3 \pm 20.8	24.4 \pm 21.0
Effectiveness/Efficiency	WSR (sym/min)	5.5 \pm 2.7	–	4.4 \pm 3.7
	Time for correct selections (s)	–	8.8 \pm 6.3	41.0 \pm 16.2
Satisfaction	SUS (0–100)	–	78.1 \pm 10.8	78 \pm 16.6

Session I, patients with WSR > 0; Session II, patients who completed the three tasks.

TABLE 4 | Scores for the NASA-tlx and SUS questionnaires in session II: Grid 3 access *via* AT input and Grid 3 access *via* BCI control.

Patient	NASA-TLX							SUS
	Workload tot	Mental demand	Physical demand	Temporal demand	Performance	Effort	Frustration	
Grid 3 access <i>via</i> AT input								
P1	4	1.3	1.3	1.3	0	0	0	95
P2	60.3	31.7	6.7	10	10	2	0	70
P3	6.7	0	2.7	1.3	0	2.7	0	87.5
P4	12.6	3.3	1.3	5.3	2	0.7	0	87.5
P5	8.3	2.7	0.3	2.7	1	1.7	0	62.5
P6	0	0	0	0	0	0	0	82.5
P7	30.3	10	0	0	2.7	2.7	15	67.5
P8	39.3	16	4.7	4.7	4	0	10	72.5
Mean ± SD	20.2 ± 21.1	8.1 ± 11	2.1 ± 2.4	3.2 ± 3.4	2.5 ± 3.4	1.2 ± 1.2	3.1 ± 5.9	78.1 ± 11.6
Grid 3 access <i>via</i> BCI control								
P1	4.7	1.3	1	1.7	0	0.7	0	90
P2	46.7	23.3	5	1.3	1.3	15	0.7	57.5
P3	40.3	10.7	4	1.3	4.7	4.7	25	62.5
P4	30.3	18.3	0	2.7	0	9.3	0	92.5
P5	48.3	18.7	2.7	24	0	1.3	1.7	72.5
P6	0	0	0	0	0	0	0	87.5
P7	29.7	13.3	0.3	3.3	6.7	2.7	3.3	75
P8	29.3	8	0	4	8	2.7	6.7	60
Mean ± SD	28.7 ± 17.9	11.7 ± 8.4	1.6 ± 2	4.8 ± 7.9	2.6 ± 3.4	4.5 ± 5.1	4.7 ± 8.5	74.7 ± 14

13 patients with MS who were admitted to the AT service of Fondazione Santa Lucia, IRCCS (Rome, Italy). Patients participated in two sessions, including (i) the control of the P300-speller (BCI control ability test; Farwell and Donchin, 1988) and (ii) the access to WhatsApp, YouTube, and Web Browser through Grid 3 first with a conventional commercial input device and then with the h-BCI system (Grid 3 access *via* AT input and *via* BCI control).

First, we tested the reliability of the newly developed h-BCI system referring to the stand-alone P300-speller, assuming the P300-speller as the most validated BCI. The comparison showed that those patients who successfully controlled the P300-speller also succeeded in mastering the h-BCI system. On the other hand, the three patients who were not able to control the h-BCI system showed similar “illiteracy” for the P300-speller control (Table 2; P5, P7, P8). Overall, these data underlined the reliability of the h-BCI system and allow us to infer that the inability to control the h-BCI system

was due to patients’ peculiarities (Oreja-Guevara et al., 2019; see below) rather than the system features. As shown by way of example in Figure 4, the amplitude of the P300 waveform over Pz for the participant who best controlled the h-BCI system (P6) was, at a visual inspection, higher with respect to the participant who did not control the system (P7). Furthermore, we noted that patients who did not control the h-BCI system showed a deficit in the executive functions, attention, and working memory (as for P5 the neuropsychological assessment was not available; Table 1). We, therefore, hypothesize that cognitive impairments could, in part, account for the inability to control the h-BCI system. This was consistent with previous studies that found a significant involvement of cognitive abilities in BCI performance in patients with amyotrophic lateral sclerosis (Riccio et al., 2013, 2018; Geronimo et al., 2016) and underlines the importance to consider the cognitive abilities in the implementation of BCI-based AT devices.

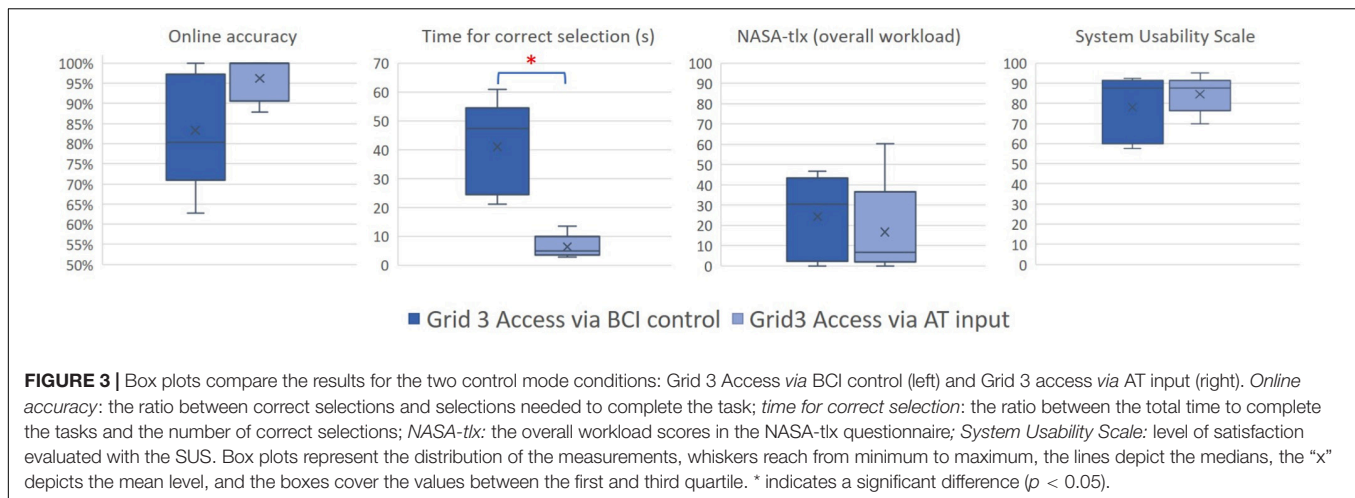


FIGURE 3 | Box plots compare the results for the two control mode conditions: Grid 3 Access via BCI control (left) and Grid 3 access via AT input (right). *Online accuracy*: the ratio between correct selections and selections needed to complete the task; *time for correct selection*: the ratio between the total time to complete the tasks and the number of correct selections; *NASA-tlx*: the overall workload scores in the NASA-tlx questionnaire; *System Usability Scale*: level of satisfaction evaluated with the SUS. Box plots represent the distribution of the measurements, whiskers reach from minimum to maximum, the lines depict the medians, the “x” depicts the mean level, and the boxes cover the values between the first and third quartile. * indicates a significant difference ($p < 0.05$).

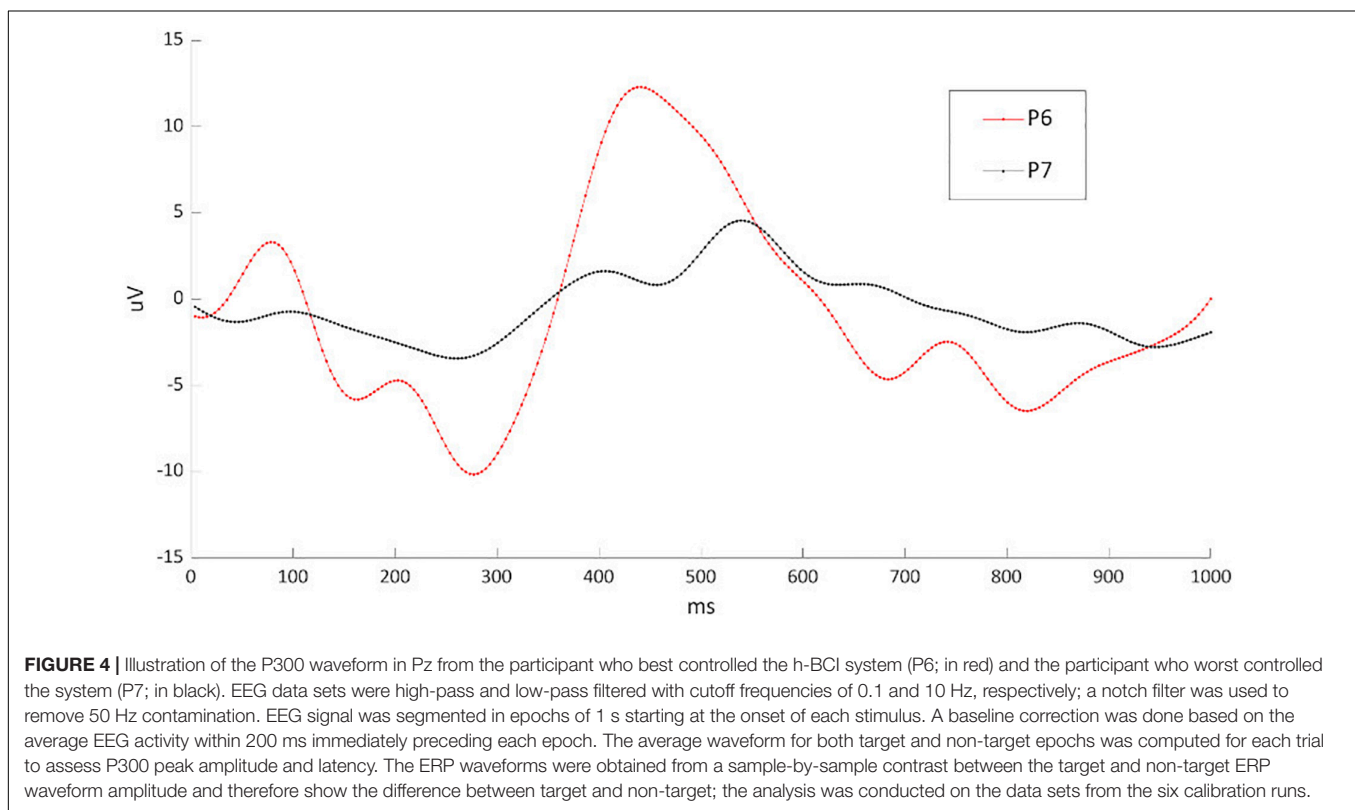


FIGURE 4 | Illustration of the P300 waveform in Pz from the participant who best controlled the h-BCI system (P6; in red) and the participant who worst controlled the system (P7; in black). EEG data sets were high-pass and low-pass filtered with cutoff frequencies of 0.1 and 10 Hz, respectively; a notch filter was used to remove 50 Hz contamination. EEG signal was segmented in epochs of 1 s starting at the onset of each stimulus. A baseline correction was done based on the average EEG activity within 200 ms immediately preceding each epoch. The average waveform for both target and non-target epochs was computed for each trial to assess P300 peak amplitude and latency. The ERP waveforms were obtained from a sample-by-sample contrast between the target and non-target ERP waveform amplitude and therefore show the difference between target and non-target; the analysis was conducted on the data sets from the six calibration runs.

Patients with MS have been rarely considered as potential BCI end-users. To the best of our knowledge, only one previous study (Martinez-Cagigal et al., 2016) reported the evaluation of a P300-based BCI (for the web browser access) with 16 patients with MS. The authors reported a control accuracy of $84.14(\pm 10.08)\%$ which is comparable with our results ($83.3\% \pm 14.6\%$ accuracy), with three patients who had a classification accuracy in the calibration session of $<70\%$, and were then excluded from the assessment.

In addition to the investigation of the ability to control the system in patients with MS, we evaluated the h-BCI system

usability according to the UCD metrics (Kübler et al., 2014; Riccio et al., 2015; Schettini et al., 2015). The comparison between the two conditions (Grid 3 access via AT input and BCI control) did not reveal significant differences both in terms of effectiveness and users' satisfaction. As for the efficiency, the access to Grid 3 was faster when conventional input devices were used compared with the BCI input. However, this disadvantage (in terms of time resources involved by the user) in the efficiency of the h-BCI system was not confirmed by the workload as perceived by patients with MS (NASA-tlx), which was comparable between the two conditions. Also, participants' satisfaction was not affected

in the h-BCI system despite the h-BCI system time demand. Thus, we can infer that taking into account the (well-known) high time demand of the BCI channel, the BCI addition as a control channel of the AT software does not worsen the usability of the whole system. However, we must consider that these results could be influenced by an effect of the fascination with high technologies associated with the fact that patients performed brief tasks proposed only once. Further developments aimed at addressing the time demand issue and consequently at improving the system efficiency would include the asynchronous control (dynamic stopping of the stimulation) and automatic control suspension features together with the automatic recalibration of the classifier's parameters (Schettini et al., 2014; Guy et al., 2018). Also, such developments would concern the evaluation of classifiers potentially more accurate than the SWLDA and the evaluation of their feasibility with the proposed system; e.g., BLDA classifier, (non)linear SVM (Manyakov et al., 2011), and Riemannian classifiers (Delgado et al., 2020). Moreover, the fact that our results did not show any relation between the system control and the patients' level of fatigue underlines the importance of deeply investigating such an issue to hypothesize an advantage of the h-BCI system to access ATs.

Although cautious due to the limited sample size and the fact that the study was conducted in an experimental setting, these preliminary findings support the reliability of a P300-based BCI as an additional input channel to access a commercial AT and the evidence that such an additional channel has no additional costs on users' perception of usability with respect to muscular-based aids. Future studies involving a larger cohort of patients should be performed to improve the power of the statistical analysis and to better investigate the potential of this (hybrid) approach in real-life scenarios. This would allow considering more variables (e.g., the comfort of the EEG cap montage, including electrodes characteristics, the use of the conductive gel, the essential presence of a skilled operator), and their influence on patients' perceived satisfaction. Furthermore, to overcome the possibility that a unique evaluation session would lead patients to overestimate the usability of the system, a longitudinal study in an ecological setting would be the next step.

REFERENCES

- Allison, B. Z., Kübler, A., and Jin, J. (2020). 30+ years of P300 brain-computer interfaces. *Psychophysiology* 57:e13569. doi: 10.1111/psyp.13569
- Andrich, R., Mathiassen, N.-E., Hoogerwerf, E.-J., and Gelderblom, G. J. (2013). Service delivery systems for assistive technology in Europe: an AAATE/EASTIN position paper. *Technol. Disabil.* 25, 127–146. doi: 10.3233/TAD-130381
- Bangor, A., Kortum, P., and Miller, J. (2009). *Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale* [US]. Available online at: <https://uxpajournal.org/determining-what-individual-sus-scores-mean-adding-an-adjective-rating-scale/> (accessed April 11, 2021).
- Choi, I., Rhiu, I., Lee, Y., Yun, M. H., and Nam, C. S. (2017). A systematic review of hybrid brain-computer interfaces: taxonomy and usability perspectives. *PLoS One* 12:e0176674. doi: 10.1371/journal.pone.0176674
- Compston, A., and Coles, A. (2008). Multiple sclerosis. *Lancet (Lond. Engl.)* 372, 1502–1517.
- Delgado, J. M. C., Achanncaray, D., Villota, E. R., and Chevallier, S. (2020). Riemann-based algorithms assessment for single- and multiple-trial P300 classification in non-optimal environments. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 2754–2761. doi: 10.1109/TNSRE.2020.3043418
- 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
- Furdea, A., Halder, S., Krusienski, D. J., Bross, D., Nijboer, F., Birbaumer, N., et al. (2009). An auditory oddball (P300) spelling system for brain-computer interfaces. *Psychophysiology* 46, 617–625. doi: 10.1111/j.1469-8986.2008.00783.x
- Geronimo, A., Simmons, Z., and Schiff, S. J. (2016). Performance predictors of brain-computer interfaces in patients with amyotrophic lateral sclerosis. *J. Neural Eng.* 13:026002. doi: 10.1088/1741-2560/13/2/026002
- Guy, V., Soriani, M.-H., Bruno, M., Papadopoulou, 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
- Hadjimichael, O., Vollmer, T., and Oleen-Burkey, M. (2008). Fatigue characteristics in multiple sclerosis: the North American research committee on multiple

Finally, we consider the integration of the BCI with the daily/commercial AT devices and the involvement of AT services in the development of innovative devices and in their customization and validation as an important step for the BCI inclusion in AT services portfolio of solutions.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Local Ethical Committee of Fondazione Santa Lucia, IRCCS. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AR was responsible for the experimental design, data collection, analysis of data, and manuscript writing. FS was responsible for the experimental design, h-BCI system implementation, analysis of data, and manuscript writing. VG contributed to participants' enrollment, data collection and analysis, and manuscript writing. EG contributed to participants' enrollment and data collection. MG contributed to participants' enrollment. DM and FC supervised the overall experimental design implementation, data interpretation, and manuscript editing. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by “Premio Merck in Neurology 2018” (Grant agreement MERCK.7; Sponsor: Merck Serono-SPA).

- sclerosis (NARCOMS) survey. *Health Qual. Life Outcomes* 6:100. doi: 10.1186/1477-7525-6-100
- Hart, S. G. (2006). Nasa-task load index (NASA-TLX); 20 years later. *Proc. Hum. Factors Ergon. Soc. Annu. Meeting* 50, 904–908. doi: 10.1177/154193120605000909
- ISO (2019). *Ergonomics of Human-System Interaction Part 210: Human-Centred Design for Interactive Systems. Multiple. Distributed Through American National Standards Institute*. Available online at: <https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/07/75/77520.html> (accessed April 11, 2021).
- Krupp, L. B., LaRocca, N. G., Muir-Nash, J., and Steinberg, A. D. (1989). The fatigue severity scale: application to patients with multiple sclerosis and systemic lupus Erythematosus. *Arch. Neurol.* 46, 1121–1123. doi: 10.1001/archneur.1989.00520460115022
- Krusiński, 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, 299–305. doi: 10.1088/1741-2560/3/4/007
- Kübler, A., Holz, E. M., Riccio, A., Zickler, C., Kaufmann, T., Kleih, S. C., et al. (2014). The user-centered design as novel perspective for evaluating the usability of BCI-controlled applications. *PLoS One* 9:e112392. doi: 10.1371/journal.pone.0112392
- Kurtzke, J. F. (1983). Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 33:1444. doi: 10.1212/WNL.33.11.1444
- Liberati, G., Pizzimenti, A., Simone, 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
- Manyakov, N. V., Chumerin, N., Combaz, A., and Van Hulle, M. M. (2011). Comparison of classification methods for P300 brain-computer interface on disabled subjects. *Comput. Intell. Neurosci.* 2011:519868. doi: 10.1155/2011/519868
- Martinez-Cagigal, V., Gomez-Pilar, J., Alvarez, D., and Hornero, R. (2016). An asynchronous P300-based brain-computer interface web browser for severely disabled people. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 1332–1342. doi: 10.1109/TNSRE.2016.2623381
- McCane, L. M., Sellers, E. W., McFarland, D. J., Mak, J. N., Carmack, C. S., Zeilind, D., et al. (2014). Brain-computer interface (BCI) evaluation in people with amyotrophic lateral sclerosis. *Amyotroph. Lateral Sclerosis Frontotemporal Degener.* 15, 207–215. doi: 10.3109/21678421.2013.865750
- Medina-Julí, M. T., Fernández-Rodríguez, Á., Velasco-Álvarez, F., and Ron-Angevin, R. (2020). P300-based brain-computer interface speller: usability evaluation of three speller sizes by severely motor-disabled patients. *Front. Hum. Neurosci.* 14:583358. doi: 10.3389/fnhum.2020.583358
- Millán, J. D. R., Rupp, R., Mueller-Putz, G., 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
- MSIF (2020). *Atlas of MS*, 3rd Edn. Available online at: <https://www.msif.org/wp-content/uploads/2020/10/Atlas-3rd-Edition-Epidemiology-report-EN-updated-30-9-20.pdf> (accessed April 11, 2021).
- Nijboer, F., Sellers, E. W., Mellinger, J., Jordan, M. A., Matuz, T., Furdea, A., et al. (2008). A P300-based brain-computer interface for people with amyotrophic lateral sclerosis. *Clin. Neurophysiol.* 119, 1909–1916. doi: 10.1016/j.clinph.2008.03.034
- Oh, J., Vidal-Jordana, A., and Montalban, X. (2018). Multiple sclerosis: clinical aspects. *Curr. Opin. Neurol.* 31, 752–759. doi: 10.1097/WCO.0000000000000622
- Oreja-Guevara, C., Ayuso Blanco, T., Brieva Ruiz, L., Hernández Pérez, M. Á., Meca-Lallana, V., and Ramío-Torrentà, L. (2019). Cognitive dysfunctions and assessments in multiple sclerosis. *Front. Neurol.* 10:581. doi: 10.3389/fneur.2019.00581
- Powers, J. C., Bieliaieva, K., Wu, S., and Nam, C. S. (2015). The human factors and ergonomics of P300-based brain-computer interfaces. *Brain Sci.* 5, 318–354. doi: 10.3390/brainsci5030318
- Rezeika, A., Benda, M., Stawicki, P., Gembler, F., Saboor, A., and Volosyak, I. (2018). Brain-computer interface spellers: a review. *Brain Sci.* 8:57. doi: 10.3390/brainsci8040057
- Riccio, A., Holz, E. M., Aricò, P., Leotta, F., Aloise, F., Desideri, L., et al. (2015). Hybrid P300-based brain-computer interface to improve usability for people with severe motor disability: electromyographic signals for error correction during a spelling task. *Arch. Phys. Med. Rehabil.* 96, S54–S61. doi: 10.1016/j.apmr.2014.05.029
- Riccio, A., Leotta, F., Bianchi, L., Aloise, F., Zickler, C., Hoogerwerf, E. J., et al. (2011). Workload measurement in a communication application operated through a P300-based brain-computer interface. *J. Neural Eng.* 8:025028. doi: 10.1088/1741-2560/8/2/025028
- Riccio, A., Pichiorri, F., Schettini, F., Toppi, J., Riseti, M., Formisano, R., et al. (2016). Interfacing brain with computer to improve communication and rehabilitation after brain damage. *Prog. Brain Res.* 228, 357–387. doi: 10.1016/b.pbr.2016.04.018
- Riccio, A., Schettini, F., Simone, L., Pizzimenti, A., Inghilleri, M., Olivetti-Belardinelli, M., et al. (2018). On the relationship between attention processing and P300-based brain computer interface control in amyotrophic lateral sclerosis. *Front. Hum. Neurosci.* 12:165. doi: 10.3389/fnhum.2018.00165
- Riccio, A., Simone, L., Schettini, F., Pizzimenti, A., Inghilleri, M., Olivetti-Belardinelli, M., et al. (2013). Attention and P300-based BCI performance in people with amyotrophic lateral sclerosis. *Front. Hum. Neurosci.* 7:732. doi: 10.3389/fnhum.2013.00732
- Rottoli, M., La Gioia, S., Frigeni, B., and Barcella, V. (2017). Pathophysiology, assessment and management of multiple sclerosis fatigue: an update. *Expert Rev. Neurotherap.* 17, 373–379. doi: 10.1080/14737175.2017.1247695
- Schalk, G., McFarland, D. J., Hinterberger, T., Birbaumer, N., and Wolpaw, J. R. (2004). BCI2000: a general-purpose brain-computer interface (BCI) system. *IEEE Trans. Biomed. Eng.* 51, 1034–1043. doi: 10.1109/TBME.2004.827072
- Schettini, F., Aloise, F., Aricò, P., Salinari, S., Mattia, D., and Cincotti, F. (2014). Self-calibration algorithm in an asynchronous P300-based brain-computer interface. *J. Neural Eng.* 11:035004. doi: 10.1088/1741-2560/11/3/035004
- Schettini, F., Riccio, A., Simone, 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, S46–S53. doi: 10.1016/j.apmr.2014.05.027
- Schreuder, M., Riccio, A., Riseti, M., Dähne, S., Ramsay, A., Williamson, J., et al. (2013). User-centered design in brain-computer interfaces—A case study. *Artif. Intell. Med.* 59, 71–80. doi: 10.1016/j.artmed.2013.07.005
- Sellers, E. W., Vaughan, T. M., and Wolpaw, J. R. (2010). A brain-computer interface for long-term independent home use. *Amyotroph. Lateral Scler.* 11, 449–455. doi: 10.3109/17482961003777470
- Smartbox Assistive Technology (2021). *Smartbox Assistive Technology thinksmartbox.com*. Available online at: <https://thinksmartbox.com/product/grid-3/> (accessed April 28, 2021).
- Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., et al. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol.* 17, 162–173.
- Thompson, D. E., Gruis, K. L., and Huggins, J. E. (2014). A plug-and-play brain-computer interface to operate commercial assistive technology. *Disabil. Rehabil. Assist. Technol.* 9, 144–150. doi: 10.3109/17483107.2013.785036
- Townsend, G., LaPallo, B. K., Boulay, C. B., Krusiński, D. J., Frye, G. E., Hauser, C. K., et al. (2010). A novel P300-based brain-computer interface stimulus presentation paradigm: moving beyond rows and columns. *Clin. Neurophysiol.* 121, 1109–1120. doi: 10.1016/j.clinph.2010.01.030
- Treder, M., 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
- Tur, C. (2016). Fatigue management in multiple sclerosis. *Curr. Treat. Options Neurol.* 18:26. doi: 10.1007/s11940-016-0411-8
- Wessels, R., Persson, J., Lorentsen, Ø., Andrich, R., Ferrario, M., Oortwijn, W., et al. (2002). IPPA: individually prioritised problem assessment. *Technol. Disabil.* 14, 141–145. doi: 10.3233/TAD-2002-14310

- Wolpaw, J. R., Bedlack, R. S., Reda, D. J., Ringer, R. J., Banks, P. G., Vaughan, T. M., et al. (2018). Independent home use of a brain-computer interface by people with amyotrophic lateral sclerosis. *Neurology* 91, e258–e267. doi: 10.1212/WNL.0000000000005812
- Zickler, C., Riccio, A., Leotta, F., Hillian-Tress, S., Halder, S., Holz, E., et al. (2011). A brain-computer interface as input channel for a standard assistive technology software. *Clin. EEG Neurosci.* 42, 236–244. doi: 10.1177/155005941104200409

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.

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.

Copyright © 2022 Riccio, Schettini, Galiotta, Giraldi, Grasso, Cincotti and Mattia. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Evaluation of a New Lightweight EEG Technology for Translational Applications of Passive Brain-Computer Interfaces

Nicolina Sciaraffa^{1*}, Gianluca Di Flumeri^{1,2}, Daniele Germano¹, Andrea Giorgi^{1,2}, Antonio Di Florio¹, Gianluca Borghini^{1,2}, Alessia Vozzi^{1,2}, Vincenzo Ronca^{1,2}, Fabio Babiloni^{1,2,3} and Pietro Aricò^{1,2,4}

¹ BrainSigns Srl, Rome, Italy, ² Department of Molecular Medicine, Sapienza University of Rome, Rome, Italy, ³ College of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, China, ⁴ Department of Computer, Control, and Management Engineering "Antonio Ruberti", Sapienza University of Rome, Rome, Italy

OPEN ACCESS

Edited by:

Hubert Cecotti,
California State University, Fresno,
United States

Reviewed by:

Ivan Volosyak,
Rhine-Waal University of Applied
Sciences, Germany
Chun-Shu Wei,
National Yang Ming Chiao Tung
University, Taiwan

*Correspondence:

Nicolina Sciaraffa
nicolina.sciaraffa@brainsigns.com

Specialty section:

This article was submitted to
Brain-Computer Interfaces,
a section of the journal
Frontiers in Human Neuroscience

Received: 21 March 2022

Accepted: 21 June 2022

Published: 14 July 2022

Citation:

Sciaraffa N, Di Flumeri G, Germano D,
Giorgi A, Di Florio A, Borghini G,
Vozzi A, Ronca V, Babiloni F and
Aricò P (2022) Evaluation of a New
Lightweight EEG Technology for
Translational Applications of Passive
Brain-Computer Interfaces.
Front. Hum. Neurosci. 16:901387.
doi: 10.3389/fnhum.2022.901387

Technologies like passive brain-computer interfaces (BCI) can enhance human-machine interaction. Anyhow, there are still shortcomings in terms of easiness of use, reliability, and generalizability that prevent passive-BCI from entering real-life situations. The current work aimed to technologically and methodologically design a new gel-free passive-BCI system for out-of-the-lab employment. The choice of the water-based electrodes and the design of a new lightweight headset met the need for easy-to-wear, comfortable, and highly acceptable technology. The proposed system showed high reliability in both laboratory and realistic settings, performing not significantly different from the gold standard based on gel electrodes. In both cases, the proposed system allowed effective discrimination ($AUC > 0.9$) between low and high levels of workload, vigilance, and stress even for high temporal resolution (< 10 s). Finally, the generalizability of the proposed system has been tested through a cross-task calibration. The system calibrated with the data recorded during the laboratory tasks was able to discriminate the targeted human factors during the realistic task reaching AUC values higher than 0.8 at 40 s of temporal resolution in case of vigilance and workload, and 20 s of temporal resolution for the stress monitoring. These results pave the way for ecologic use of the system, where calibration data of the realistic task are difficult to obtain.

Keywords: passive-BCI, EEG, water-based electrodes, workload, vigilance, stress, human factors

INTRODUCTION

The human-machine interaction involves all the processes of interaction and communication between a human user and a machine. It aims to make the human even more "understandable to the machine" by systems that can detect and classify human feelings, emotions, and cognitive states (Di Nardo et al., 2020). In recent times, more and more attention arose around the use of brain signals to detect the human state, that is the passive brain-computer interfaces (BCI). BCI was born for clinical needs and progressively moved from translating intentional (actions) to unintentional (mental states) brain outputs, thus, from active to passive-BCI (Zander et al., 2009). The passive-BCI does not aim to voluntarily control the external world but to allow the surroundings to automatically adapt their behavior to the actual subject's mental states

(Zander and Kothe, 2011; Aricò et al., 2018). The impact of this new aspect of the BCI is clear for safety-critical environments, where the mental state of users is usually correlated to the probability of making errors or, more generally, to human factors (Aricò et al., 2017b; Zander et al., 2017). The monitoring of the user-state is just one of the countless applications of passive-BCI that have been resumed in 5 categories. The additions to user-state monitoring are safety and security, training, education, cognitive improvement, gaming, and entertainment (Erp et al., 2012).

Despite the growing interest in passive-BCI applications, there is still a gap between theory and practice. On the one hand, the advantage of using a human-centered design is clear, since it could facilitate the operator's work and reduce safety risks (Ke et al., 2021). On the other hand, ethical and technical limitations prevent the growth of passive-BCI technology within operational contexts: ease of use, reliability, and generalizability are the key characteristics affecting the end-users perception and choice of a passive-BCI system (Douibi et al., 2021).

Technical Limitations of Passive-BCI Systems

In this context, electroencephalography (EEG) is the most used method to measure the electrical activity of the brain. It is usually preferred, thanks to its portability, non-invasiveness, and high temporal resolution (Roman-Gonzalez, 2012). However, a traditional passive-BCI setup requires the use of Ag/AgCl electrodes that are in contact with the scalp through the electrolytic gel. This strongly limits the ease of use (Webster, 2009). In fact, in place of excellent signal quality, these recording systems required the scrubbing of the skin to lower the electrode impedance before using electrolytic gel. This is a time-consuming activity that takes 20 min or more depending on the number of electrodes used. Extra time is also needed because the subject should wash the hair to remove gel residues. This approach and gel-based electrodes are still considered the gold standard. However, the gap between the new water and gel electrodes is narrowing (Volosyak et al., 2010; Chi et al., 2012; Lopez-Gordo et al., 2014; Di Flumeri et al., 2019a). Water-based electrodes consisted of felt or similar tissue soaked in water or saline solution to connect the electrode to the skin (Müller-Putz and Wriessnegger, 2021). This makes both setup and clean-up faster and easier, as well as more comfortable without affecting the quality of the signal (Volosyak et al., 2010). In this regard, different portable and gel-free EEG recording systems have been employed in several studies: the tested electrodes were able to guarantee the same quality levels as the gel electrodes, along with significantly faster setup and higher comfort (Volosyak et al., 2010; Amaral et al., 2017; Vourvopoulos et al., 2017; Di Flumeri et al., 2019a).

Despite this, the wearable EEG devices available on the market are a step behind other wearable devices. A recent analysis of the current market found 28 companies developing wired and wireless EEG equipment (Jamil et al., 2021). However, it has not yet been possible to converge comfort with pleasant

appearance, effective recording, and adaptability to different operating contexts.

Despite the wide range of recording systems, the time-to-market estimate made 10 years ago for passive-BCI systems has not yet been met (Erp et al., 2012). This is also because the reliability of a passive-BCI system depends not only on hardware but also on software aspects. On the one hand, the recording technology should guarantee ease of use without adversely affecting signal quality. On the other hand, the passive-BCI system should be accurate and not bore the user with frequent calibrations. This aspect depends on the suitability of the algorithm implemented in the system. First, signal processing techniques should be adapted for out-of-the-lab use. For example, the low number of channels expected for wearable EEG and the need for high reactivity disfavor the most widely used signal correction methods, such as Independent Component Analysis (Makeig et al., 1995). Furthermore, the more the number of pre-processing steps, the higher the computational effort required, resulting in larger computational modules and faster battery consumption. In addition, for online use, all methods that need to know the entire data (e.g., normalization methods) cannot be used. Once clean EEG signals have been obtained, further steps have to be performed to obtain features that can identify the user's mental state. Spectral features can be efficiently calculated and can provide real-time feedback. Such features are usually used to train machine learning or deep learning algorithms (Aricò et al., 2016). These methods must be reliable, but must not require frequent calibration, ideally zeroing it out with an unsupervised approach (Schultze-Kraft et al., 2016). Although unsupervised approaches may allow greater generalizability of the system, supervised methods are preferred as they allow significantly higher accuracies to be achieved (Blankertz et al., 2016). Better generalization results can be achieved using a cross-task approach. An open issue related to the use of passive-BCI during a realistic task is the lack of specific calibration tasks able to elicit a specific mental state in the subject. Well-designed laboratory tasks can be effective for calibration use in a cross-task approach.

To summarize, EEG technology and calibration of the passive-BCI system remain the two main gaps for out-of-the-lab use.

Targeted Human Factors

Drivers, air traffic controllers, surgeons, athletes, and pilots represent some of the users who could benefit most from the use of passive-BCI systems during both training and operational conditions (Aricò et al., 2017a; Zander et al., 2017; Alimardani and Hiraki, 2020). In general, the common denominator of these application areas is human performance: passive-BCI could allow obtaining information on the user's psychophysiological state, in terms of different human factors, to develop solutions for improving performance and safety. Building on this, three human factors have been targeted in the current work: mental workload, stress, and vigilance.

The mental workload has been chosen due to its relevant role in safety-critical applications (Young et al., 2015). In general, the mental workload quantifies the mental resources used during a task and is correlated with the performance

(Wickens, 2008). Extremely high (i.e., overload), while extremely low (i.e., underload) workloads have been associated with poorer performances. Therefore, a passive BCI system could intervene to detect a fatal overload/underload condition. Assessment of mental workload by EEG provides a reliable and objective measure that can be used for this purpose in different laboratory and realistic contexts (Zhou et al., 2021).

Stress is another relevant human factor in many operational domains (Aricò et al., 2017b). One of the definitions of stress is “a state that occurs when demand outstrips coping strategies” (Hobfoll and Shirom, 1993). High task demand, frustration, time pressure, uncontrollability, and negative judgment are typical stressors that adversely affect performance by altering the cognitive processes underlying decision-making, attention, and memory (Skoluda et al., 2015). Stress assessment by EEG takes advantage of the higher temporal resolution and the possibility of direct access to the activity of the central nervous system, measuring stress levels continuously and without interfering with human activities (Borghini et al., 2020). Moreover, there are several examples of real-time measures of stress based on EEG signals using machine learning methods (Saeed et al., 2015; Attallah, 2020; Halim and Rehan, 2020; Hekmatmanesh et al., 2021; Katmah et al., 2021).

Finally, monitoring levels of vigilance has proven to be a key aspect in several fundamental environments (Neigel et al., 2020). The concept of vigilance belongs to the broader domain of attention. A physiological decrease in vigilance over time is associated with performance degradation, such as slower reaction times and loss of situation awareness, whereas optimal performance is ensured by an adequate level of activation throughout the task (Parasuraman et al., 1998). EEG measures have been used as features for machine learning models to monitor vigilance levels in different contexts (Sebastiani et al., 2020; Kamrud et al., 2021; Li and Chung, 2022).

In this framework, the present work aims to evaluate a new lightweight EEG technology for translational use of passive-BCI. In particular, the system based on water-based electrodes was tested in terms of impedances, artifacts, usability, and human factors assessment and was compared with a gel-based equivalent that was used as a gold standard. To analyze the reliability and generalizability of the proposed system, it has been tested during two studies focusing on laboratory and realistic settings, respectively.

MATERIALS AND METHODS

Experimental Participants

Twenty healthy subjects (29.3 ± 5.12 years old) took part in the study: 8 women and 12 men were recruited voluntarily. Informed consent was obtained from each subject and all the data were pseudorandomized to prevent any association with the subject's identity. The experiments were conducted following the principles outlined in the 1975 Declaration of Helsinki, as revised in 2000. Experiments were approved by the Ethical Committee of the Sapienza University of Rome. Due to the involvement in a realistic driving experiment, a driver's license and the ability

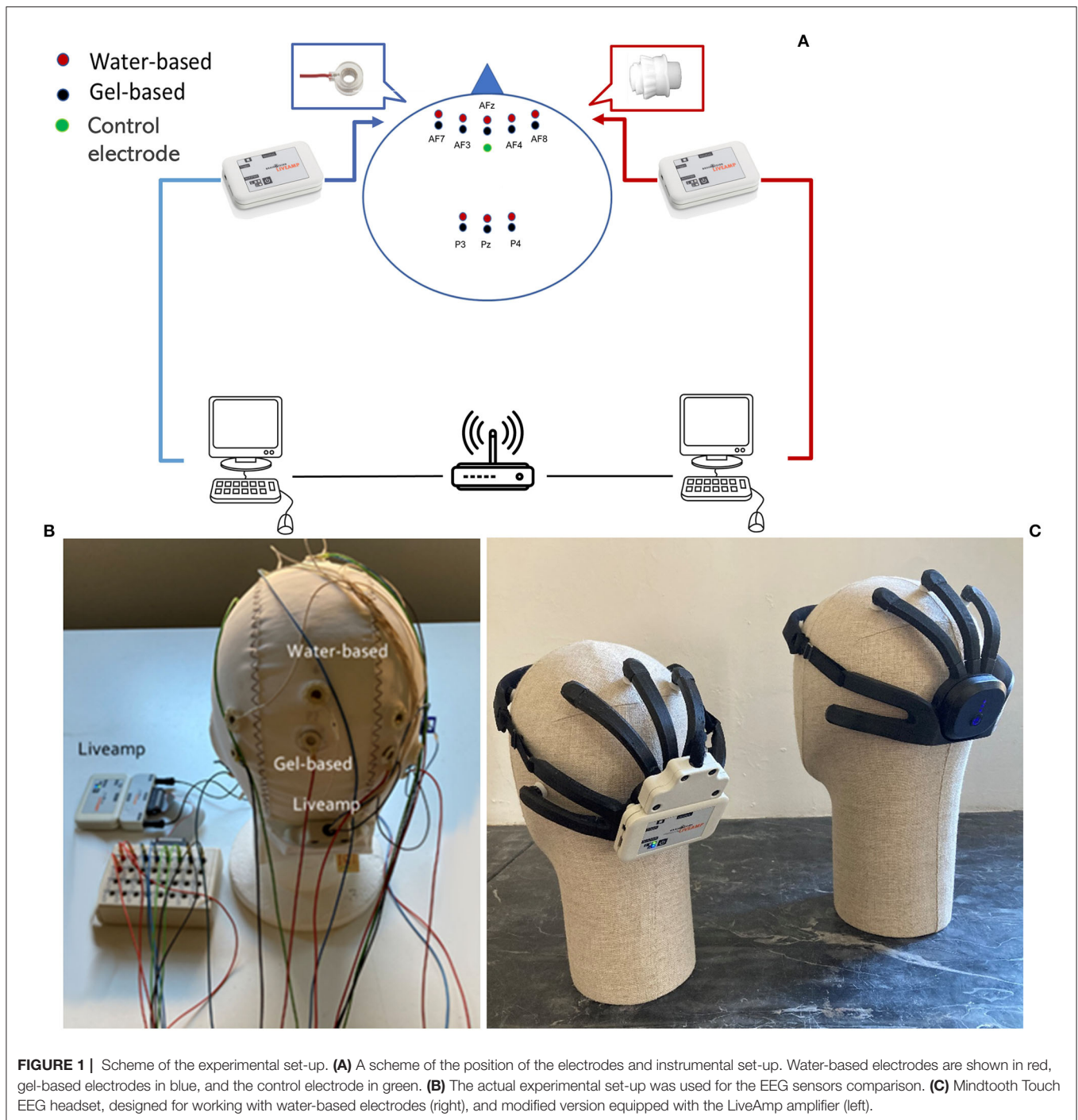
to drive a car with a manual transmission were required as inclusion criteria.

Experimental Set-Up

Gel and water-based electrodes were used for this experiment (Figure 1A). The gel-based electrodes were traditional Ag/AgCl electrodes produced by EasyCap GmbH (Woerthsee-Ettersschlag, Germany). The water-based electrodes consisted of open-celled, hydrophilic, and highly absorbent cylindrical sponges produced by Brain Products GmbH (Gilching, Germany). It hardens when dry and becomes soft and expandable when wet. The porous material is intended to soak aqueous electrolyte solutions (1–2% sodium chloride solutions are used as electrolytes). The electrolyte solution is used to accomplish an easy electrically conducting connection to the skin of the subject. While this mechanism of aqueous electrolyte is similar to the one of gel electrodes, the big advantage of aqueous electrolyte is that it does not leave significant amounts of residuals in the hair; the water in the electrolyte evaporates over time just leaving tiny amounts of salt on the skin that is removed with the next hair wash. To slow down the evaporation of water, a soft pedestal hermetically sealed the chamber. The sponges are integrated into a holder that can be easily connected to the electrode base using a bayonet twist lock (i.e., sponge holder). When the sponge holder is screwed into its counterpart on the electrode, the other end of the sponge comes into contact with an electrode pellet. The electrode pellet is made of a conductive plastic that is coated with a thin layer of silver/silver chloride (Ag/AgCl), which is an electrically stable and widely used electrode material.

Both gel-based and water-based electrodes have been positioned at a 1.5 cm of distance between each other on a standard EEG cap (Figure 1B). This has been done to maximize the hypothesis to ideally record from the same point and to guarantee a minimum distance to not generate short circuits between near electrodes. The investigated scalp positions corresponded to AFz, AF3, AF4, AF7, AF8, Pz, P3, and P4 of the 10–10 International System. These specific channels have been chosen for both theoretical and practical reasons. On the one hand, these positions should guarantee the assessment of the identified human factors covering the required brain areas as known from previous evidence (please refer to “Targeted Human Factors” in the Section Introduction). On the other hand, a low number of electrodes is necessary to foster ease of use. It was also inserted a gel-based “Control” electrode, at a 1.5 cm of distance to the AFz position to quantify the possible difference between near couples of electrodes because of the distance among them.

For each set of electrodes, we used a LiveAmp amplifier (Brain Products GmbH, Gilching, Germany) (Figure 1A). Each amplifier has been connected to a homemade recording software developed in Python and based on Lab Streaming Layer (LSL) communication protocol. This software was run on two tablets with Microsoft Windows 10 operating system. The two tablets were connected to the same intranet, by using a wired LAN connection. Thanks to the LSL protocol and the LAN wired connection, it was possible to guarantee a perfect synchronization among all the connected devices, unless there is a lag of two times the inverse of the sampling rate (i.e., 8 ms).



In addition, the water-based electrodes have also been tested when embedded in the Mindtooth Touch EEG headset (referred to as “New Structure” in the Section Results) to maximize the comfort for the user, sensor contact, and easiness to fit (**Figure 1C**). In particular, the front part of the headset hosts the five anterior-frontal EEG sensors and is designed to fit the user’s forehead. To maximize comfort, it is elastic and made of a combination of flexible materials. The front part has removable

padding made in polyurethane (PU) foam to further improve comfort. The rear part hosts the ground and reference sensors, the three parietal EEG sensors, and the EEG amplifier. Ground and reference sensors are fitted to two separate “fingers” on the mastoids. The “fingers” of the three parietal EEG sensors are made as a separate part that is fitted and fixed to the rear part of the headset. All the parietal EEG sensors’ holders have been prolonged and fitted to a rotating end-piece “a fingertip”

to make it easier to adapt to all head shapes, split the hair, and touch the scalp. The hardware within the EEG amplifier has been developed by Brain Products (GmbH, Gilching, Germany). For the specific purposes of this experiment, the original amplifier of the Mindtooth Touch headset has been replaced with a LiveAmp amplifier, so as to not induce any bias due to different hardware.

EEG Acquisition and Preprocessing

Once arrived, each subject was asked to wear the EEG standard cap with both the electrode types. Sponges of water-based electrodes were first disinfected using a solution of water (70%) and isopropanol (30%) and then partially soaked with 1% of NaCl saline solution. The quantity of saline solution was limited to the needed quantity to have a good signal quality, but not too much to induce dripping, and so as to generate possible shortcuts with near gel-based electrodes. At this point, the conductive gel was applied to all the gel-based electrodes in a limited quantity, so as to not generate any shortcuts between near electrodes. Reference electrodes of both the systems were put on the earlobe, and the ground on the left mastoid, by assuring even there a minimum distance to not have any shortcuts. Before starting the recording, the impedances have been brought below 40 k Ω (Ferree et al., 2001) and 100 k Ω (Kappenman and Luck, 2010) threshold for gel and water-based sensors, respectively.

The EEG signal was first band-pass filtered with a fifth-order Butterworth filter in the interval 2–30 Hz. The blink artifacts were detected using the Reblinca method (Di Flumeri et al., 2016) and were corrected by leveraging the ocular component estimated through a multi-channel Wiener Filter (MWF) (Somers et al., 2018). EEG signals were segmented into epochs of 1 s, and the threshold criterion ($\pm 80 \mu V$) was applied for artifact rejection (Hubbard et al., 2019). This conservative value has been preferred to 100 μV , the default value suggested within the EEGlab toolbox, because just one criterium has been adopted (Delorme and Makeig, 2004).

Experimental Protocol

The experimental protocol consisted of a sequence of laboratory (Study 1) and realistic (Study 2) tasks. They have been proceeded by two rest conditions: the subjects were asked to stay 1 min with closed eyes and 1 min with open eyes looking at a white cross on a screen. The closed eyes condition has been used to compute the individual alpha frequency (IAF) value (Klimesch, 1999). Since the alpha peak is mainly prominent during rest conditions, the subjects were asked to keep their eyes closed for a minute before starting the experiment. Such a condition was then used to estimate the IAF value specifically for each subject. The open eyes condition has been used to compare the two systems in terms of spectral features (Lopez-Gordo et al., 2014).

Therefore, the subjects were asked to familiarize themselves with the tasks. During task familiarization, they wore the Mindtooth Touch headset embedded with water-based electrodes. In this time interval of about 40 min, the impedances have been evaluated and saved at the end of each sample task. These values have been compared with the impedances of water-based electrodes recorded while placed on the standard EEG cap to compare the stability of the signals against the Mindtooth

Touch Headset. Moreover, at the end of each experimental condition during both studies 1 and 2, impedance values of gel and water-based sensors were evaluated and saved to compare the stability of the impedance values between the two kinds of sensors.

Study 1: Test Laboratory Tasks

The aim of Study 1 was to analyze the performance of the light EEG system during standard controlled laboratory tasks like Multitasking and psychomotor vigilance task (PVT) as well as to provide the data for calibrating the passive-BCI system for a cross-task application.

First, the subjects were asked to accomplish the multitasking. The multitasking application comprises a set of four concurrent cognitive tasks of varying difficulty presented *via* split-screen (Figure 2A). The four chosen tasks are:

1. Mental arithmetic (left-up): The addition results must be entered into the numeric keypad. As the difficulty increases, the number of digits (from 1 to 3) and carryover digits (from 0 to 2) increases.
2. Auditory monitoring (right-up): A target tone must be identified between two tones of different frequencies emitted at regular intervals. As the difficulty increases, the target tone and distractor tone increase in similarity.
3. Visual monitoring (left-down): A horizontal fill bar should be reset as soon as it becomes full. As the difficulty increases, the fill rate increases.
4. Phone number entry task (right-down): A number must be entered on a keypad. As the difficulty increases, the number of digits to be entered increases (from 4 to 10).

According to the literature, performing several concurrent tasks compared to the single-task approach induces an increase in mental workload (Comstock, 1992; Wetherell and Sidgreaves, 2005). In addition, increased difficulty induces a greater perception of time pressure and frustration, resulting in increased stress. Therefore, the subjects were asked to perform the four tasks individually. In this case, a 30-s condition was performed for the auditory monitoring, visual monitoring, phone number entry task, easy mental arithmetic task, and hard mental arithmetic task. After that, subjects performed the four concurrent tasks (multitasking phase) at the same time. In this regard, 7 levels of 1 min multitasking at increasing difficulty have been performed. To confirm workload and stress manipulation, performance measures have been recorded. We hypothesized that the subject experienced the following: (i) low mental workload during the single task execution; (ii) high mental workload during the multitasking execution; and (iii) increasing stress during the 7 levels of multitasking at increasing difficulty.

At the end of multitasking, the subjects were asked to perform the PVT to elicit different levels of vigilance. The PVT is a computerized version of the Wilkinson and Houghton task (Wilkinson and Houghton, 1982) aiming to analyze the decrease of vigilance in 10 min (Figure 2B). During this task, subjects are asked to press the space bar on the keyboard as quickly as possible in response to a red circle appearing on the screen for 1 s after a fixation cross. The inter-stimulus intervals range

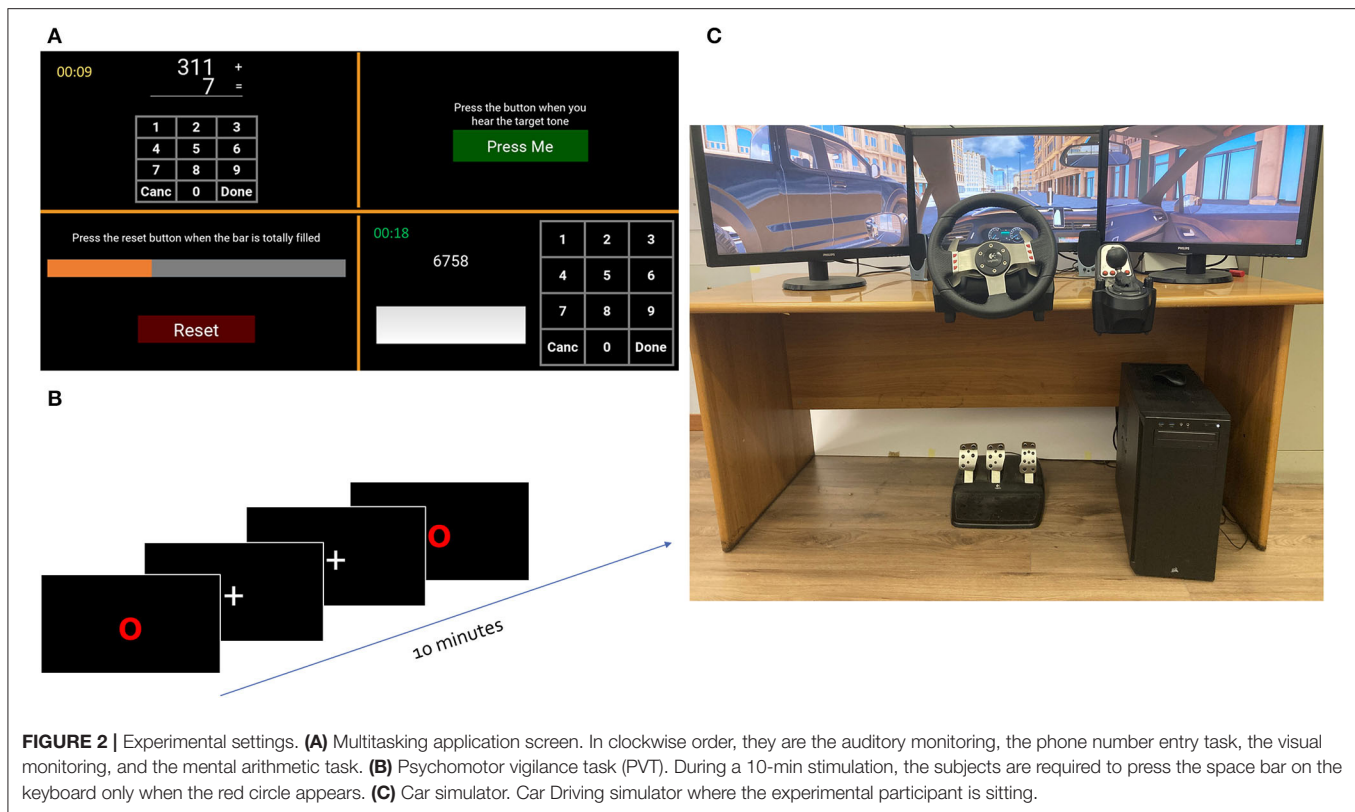


FIGURE 2 | Experimental settings. **(A)** Multitasking application screen. In clockwise order, they are the auditory monitoring, the phone number entry task, the visual monitoring, and the mental arithmetic task. **(B)** Psychomotor vigilance task (PVT). During a 10-min stimulation, the subjects are required to press the space bar on the keyboard only when the red circle appears. **(C)** Car simulator. Car Driving simulator where the experimental participant is sitting.

randomly from 2 to 5 s. According to the literature (Molina et al., 2019), this monotonous task can induce a decrement of vigilance in participants, therefore we hypothesized that the experimental sample experienced: (i) a high level of vigilance during the first 2 mins (levels PVT1 and PVT2); (ii) low level of vigilance during the last 2 mins (levels PVT9 and PVT10).

Study 2: Test Realistic Driving Task

The aim of Study 2 was to analyze the performance of the light EEG system during realistic driving and to test the passive-BCI system calibrated with data recorded during the laboratory tasks. The car simulator is a typical real car setup (seat and dashboard with the steering wheel, gearshift, and all the common commands), while the virtual environment is reproduced through three 27" screens (**Figure 2C**). This configuration was enough to make the driving environment immersive for the participants, thanks to the high realism of the driving environment. The implemented exercise allowed the elicitation of variations in workload, stress, and vigilance. The task for the participants was to follow the navigator instructions shown on the central screen.

In particular, for the mental workload assessment, two repetitions of easy and hard exercises have been generated. In the first repetition, the exercise was set in a circuit, and in the second repetition, it was set in an urban context with traffic. The level of difficulty was changed by acting on the number of curves,

intersections, bottlenecks, pedestrians crossing the street, and the number of cars. Easy and hard conditions have been randomized among subjects.

For the stress scenarios, participants were instructed to perform the same two-run repetitions just completed, but with three more stressors:

1. Time pressure: Participants have a limited time to make the same route. A chronometer was used to show the remaining time.
2. White coat effect: An operator was just behind the driver, taking notes of errors committed.
3. External noise: Heavy urban traffic noise has been presented during the exercise.

Finally, the subjects were asked to complete the scenario used to induce decrement in vigilance. It was a 10-mins drive on the highway, without any traffic, in which the participant had to drive all the time in first gear in complete silence (i.e., the sound of the car has been muted).

The car simulator was able to generate a log file with all the specific events that could happen within each scenario execution. Among the possible available events, collisions of the car with anything in the simulated world have been considered performance indicators. In particular, the more the number of collisions, the more difficult or stressful the specific level was. We analyzed the number of collisions for the easy, hard, and stressful conditions.

Performed Analysis

Taking into account the spatial and temporal variability of the EEG signal, standard procedures to compare two or more recording systems make use of a comparison to a gold standard (i.e., gel electrodes) through two possible approaches, either simultaneously or serially (Ruffini et al., 2008). Most encountered parameters for evaluating EEG signal quality are comparisons of the power spectral densities, analysis of signal-to-noise ratio (SNR), and test of electrodes during applications (Tăutan et al., 2014).

Signal Quality Analysis

In this work, the two systems were used for a simultaneous recording of EEG signals from contiguous scalp positions. Since our final aim is to use the proposed system for passive-BCI applications, for which spectral features will be used, the two systems have been compared in terms of variation of the impedance values over time and power spectra correlation.

Gel and water-based electrode impedance were considered good enough, for values respectively below 40 k Ω (Ferree et al., 2001) and 100 k Ω (Kappenman and Luck, 2010). Since the acceptability thresholds for the starting impedance values were set at two different values according to the literature, the impedances were not compared between the two systems for the same time point. However, to analyze the stability of each system during the recording, the impedances of the analyzed systems were evaluated at the beginning and the end of each experimental condition, and then, were linearly interpolated to have an impedance value every 10-min during a 40-min interval. Similarly, the effects related to the use of an unoptimized structure as sensor holders have been analyzed. Leveraging the time spent by the subject for task familiarization, they were asked to wear the Mindtooth Touch headset and, also in this case, the impedances were evaluated at the beginning and the end of each sample condition, and then, were linearly interpolated to have an impedance value every 10 min during a 40-min interval. It is noteworthy to highlight that in this case there is a temporal delay in impedance assessment because they have been measured in two different moments. The Wilcoxon signed-rank test (Bonferroni corrected) was performed to test the difference of each electrode concerning the starting point.

As impedance values affect SNR, we compared the number and the distribution of artifacts of the two systems. The Wilcoxon signed-rank test was performed to test the difference between the two systems in terms of the number of artifacts.

Finally, the power spectral densities (PSD) obtained through the two systems during open eyes conditions for each channel have been compared in terms of correlation. The higher the correlations between the frequency spectra of the gel and water electrode, the higher the possibility that both systems are equally capable of measuring spontaneous EEG (Zander et al., 2011). The open-eyes condition has been chosen because the EEG signal can be considered stationary during this condition. For each subject, Pearson's correlation has been computed between each couple of contiguous channels (e.g., AF3-water and AF3-gel) and for each band, that are Theta (3–7 Hz), Alpha (7–13), Beta (13–26), and High Beta (21–26). Each correlation value was Bonferroni

corrected for multiple comparisons, therefore the threshold for significance has been set to $\alpha = 0.0025$ (corresponding to a correlation value of $R = 0.423$). The obtained values have been statistically compared with the values of correlation obtained correlating the “Control” electrode and the AFz-gel electrode through a Wilcoxon signed-rank test (Bonferroni corrected). As they are both gel electrodes 1.5 cm apart, the obtained correlation can be considered the maximum achievable and the difference in correlation was only due to the distance between them.

Usability

To evaluate the usability and the comfort of the new system, participants were asked to answer specific questions regarding these aspects. In particular, the usability questions investigated the easiness of putting on and taking off the Mindtooth Touch headset. Participants could choose among four answers: Easy, Acceptable, Hard, Impossible. To assess the comfort, the participant had to choose among four possible answers: Comfortable, Acceptable, Tolerable, and Uncomfortable.

Neurometrics

For each human factor (workload, stress, and vigilance) a subset of channels and bands of interest have been chosen according to the literature and previous results (see Section Introduction for details). We defined “Neurometric” as the measure of each Human Factor, by using a specific combination of Global Field Power values (Di Flumeri et al., 2018). The Global Field Power (GFP) was calculated using the clean EEG for each frequency band of interest. The bands were defined accordingly with IAF values. Consequently, the following EEG bands were defined (Klimesch, 1999):

- Theta = (IAF – 6): (IAF – 2) Hz
- Alpha = (IAF – 2): (IAF + 2) Hz
- Beta = (IAF + 2): (IAF + 16) Hz
- Beta High = (IAF + 11): (IAF + 16) Hz

The workload neurometric is defined as the ratio between frontal activity in theta and parietal activity in the alpha band (Borghini et al., 2013). The theta rhythm increases, especially over the pre-frontal cortex, when the mental workload increases, while the alpha rhythm presents an inverse correlation with the mental workload, especially over the parietal cortex (Gevins et al., 1997). Therefore, the workload neurometric in this work was defined according to the literature as:

$$\text{Workload} = \frac{\text{Theta}(\text{AF8}, \text{AF7}, \text{AFz}, \text{AF3}, \text{AF4})}{\text{Alpha}(\text{P3}, \text{P4}, \text{Pz})}$$

Regarding stress measurement, the literature proved that there is a correlation between cortisol and brain activity in beta (Seo and Lee, 2010). A stress neurometric based on parietal brain activity in the high beta band was effectively tested during laboratory multitasking and was validated during realistic driving, proving that it is possible to have reliable stress measurements using the following (Sciaraffa et al., 2022):

$$\text{Stress} = \text{BetaHigh}(\text{P3}, \text{P4})$$

From the neurophysiological point of view, vigilance-related processes involve mainly the right inferior frontal brain regions (Di Flumeri et al., 2019b; Sebastiani et al., 2020; Sciaraffa et al., 2021). Increased frontal activity in the beta band, more in right than in the left hemisphere, is correlated with vigilance decrement (Molina et al., 2013). Therefore, the vigilance neurometric was defined as:

$$\text{Vigilance} = -\text{Beta}(\text{AF4}, \text{AF8})$$

The effectiveness of each neurometric (obtained by means of the water-based and the gel-based electrodes) was tested by statistically comparing through the Wilcoxon signed-rank test, their efficiency in discriminating experimental conditions. In particular, the following are compared for both laboratory and realistic tasks:

- workload neurometric was compared between easy and hard conditions;
- stress neurometric was compared between low and high-stress conditions;
- vigilance neurometric was compared between high and low vigilance conditions.

Machine Learning Model Calibration

Once obtained the neurophysiological information characterizing each mental state of interest, a machine learning-based approach has been employed to classify the two levels of each mental state (easy vs. hard for workload, and low vs. high for stress and vigilance, respectively) even at different time resolutions. The GFP values were used in this case as features:

- workload: AF7, AF8, AF3, AF4, and AFz in the Theta band and P3, P4, and Pz in the Alpha band (8 Features);
- stress: P3 and P4 in Beta High band (2 Features);
- vigilance: AF4 and AF8 computed in the Beta band (2 Features).

Each observation belonging to a specific run was labeled according to the level of mental state expected for that run. In the case where the classes (e.g., easy and hard) were too unbalanced because of the number of rejected artifacts, the Adaptive Synthetic (ADASYN) method was used to provide synthetic resampling (He et al., 2008). Once the observations for each class were balanced, and intra-subject approach was used to train the machine learning model (i.e., both training and testing were performed using the observations coming from the same subject). At this point two different procedures were performed, depending on the number of repetitions available for each task:

1. For each task having two repetitions per condition (i.e., the workload in a realistic setting, vigilance, and stress in both laboratory and realistic setting) one repetition has been used to train the algorithm and optimize the parameters, and one to test the algorithm.
2. For each task that does not have two repetitions per condition (i.e. workload in a laboratory setting), the k-fold cross validation has been performed (Schaffer, 1993). This allows to divide the dataset in k ($k = 3$) fold and to use two of them as

a training dataset and one as a test dataset. Even in this case, all the possible combinations of training and testing sets have been analyzed.

Following preliminary analysis, the Random Forest was chosen as the model and the number of estimators and the max depth have been optimized in the range from 50 to 500, and from 1 to 50, respectively (Breiman, 2001).

The effectiveness of each model was assessed by computing the area under curve (AUC) (Bamber, 1975) between 0 and 1 at different temporal resolutions from 1 to 60 s. The whole described classification procedure has been repeated for water-based and gel-based systems and the two AUC curves obtained at different time resolutions have been statistically compared through Wilcoxon signed-rank test.

To test the generalizability of the proposed water-based passive BCI system, the same Random Forest model has been calibrated through a cross-task approach. For each human factor, the data recorded during the laboratory tasks (Study 1) has been used to calibrate the model. It has been then tested on the data recorded during the realistic driving (Study 2), and the AUC has been reported for time resolution from 1 to 60 s. This approach has been statistically compared with the intra-subject approach already described through Wilcoxon signed-rank test. Finally, each identified model was used to predict the probability of a specific observation belonging to a realistic session to have a high value of mental state (i.e., to belong to the high class). These curves were compared in terms of Pearson's correlation and distances between the curves using the root mean squared error (RMSE).

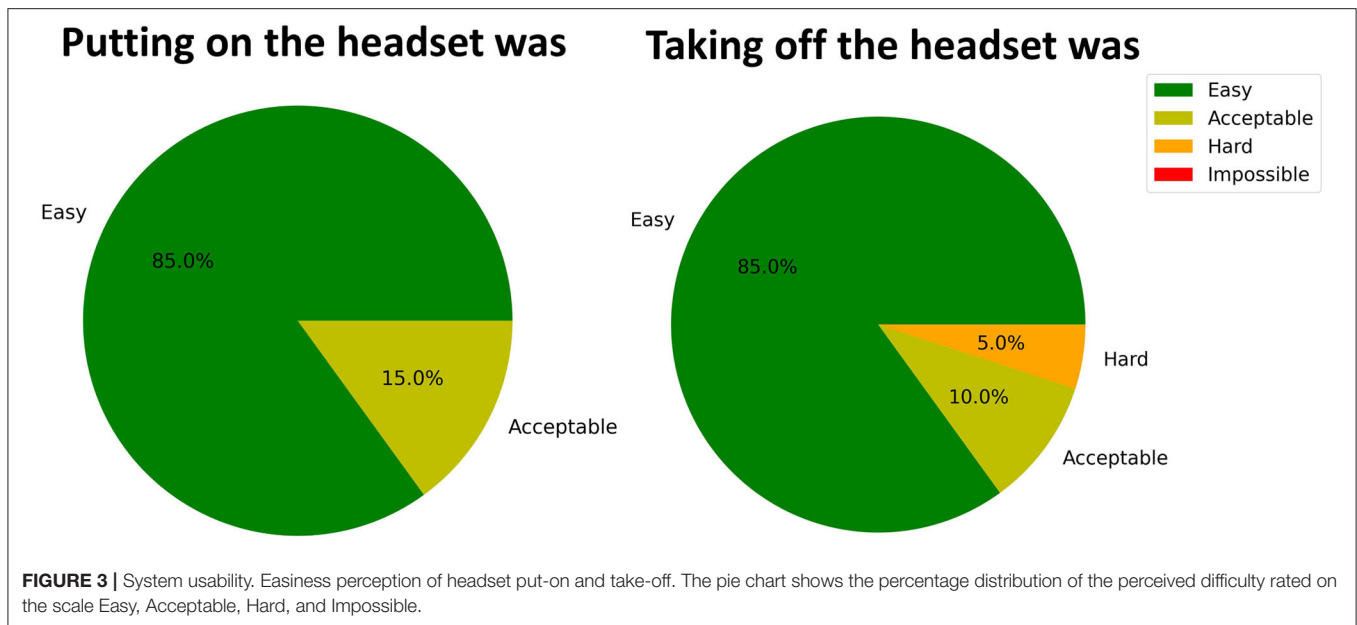
RESULTS

Usability

For the usability assessment, the first aspect considered was referred to the easiness to wear the Mindtooth Touch headset. Putting on the headset was "Easy" for the 85% and "Acceptable" for the 15%. Moreover, we found that the time needed to put on the headset was 10 s on average (Figure 3). Therefore, participants were asked to take off the headset, and rate how it was. Eighty-five percent replied that it was "Easy," 10% "Acceptable," and 1 participant "Hard." Regarding comfort, after 10 min of subjects' wearing the headset, 70% of them felt it was "Comfortable" and 30% felt it was "Acceptable".

Signal Quality

Figure 4A shows that the water electrodes exhibited significant higher impedance values compared to starting point depending on the position and the duration of the recording. In particular, AF7 ($Z = 26, p = 0.041$), AF8 ($Z = 26, p = 0.008$), and Pz ($Z = 18, p = 0.002$) showed significantly higher impedances after 10 min of recording; AF4 ($Z = 32, p = 0.019$) and P4 ($Z = 37, p = 0.038$) after 30 min. The same electrodes embedded on the Mindtooth Touch Headset exhibited lower impedances even if all the frontal electrodes significantly increase over time starting from 30 min: AFz ($Z = 30, p = 0.015$), AF3 ($Z = 34, p = 0.026$), AF4 ($Z = 17, p = 0.002$), AF7 ($Z = 19, p = 0.01$), and AF8 ($Z = 13, p = 0.0006$).



The gel electrodes did not show significant differences apart from the electrode AF7 after 40 min ($Z = 142$, $p = 0.008$).

The following analysis shows the number of artifacts for each recording technology (**Figure 4B**). The analysis shows a significantly higher number of artifacts by using the water compared to the gel sensors for frontal sites ($Z = 198$, $p = 0.0001$). Moreover, both the systems exhibited a higher number of artifacts on frontal sites compared to parietal sites (Water: $Z = 210$, $p < 10^{-5}$; Gel: $Z = 207$, $p < 10^{-5}$). Anyhow, the number of artifacts with water-based technology was about 5% on average.

At this point, we investigated if the higher impedance values of the water technology negatively affect the spectral features used to compute EEG neurometrics related to the selected human factors. **Figure 5A** shows the averaged spectra obtained with both sensors after the artifacts' rejection. These spectra were analyzed in terms of Pearson's correlation (R). **Figure 5B** shows the distributions of the R values for each frequency band. Nine of the 720 obtained R values were lower than 0.423, therefore, not significant according to Bonferroni correction. The correlation is on average higher than 0.95 for each channel and band. These distributions were statistically compared with the R values obtained correlating the spectra of AFz-gel and control electrode. The Wilcoxon signed-rank test (Bonferroni corrected) shows that the correlation between gel and water spectra is significantly lower compared to the Control-AFz spectra especially for alpha and beta bands, whereas this is less evident in the theta band.

Study 1: Test on Laboratory Tasks

During multitasking tasks, single and multitasking levels were considered as low and high workload conditions, respectively. **Figure 6A** shows that the Neurometrics of workload computed by using both the technologies showed a significant difference between the easy and the hard condition (Gel: $F = 19$, $p = 0.0006$; Water: $F = 14$, $p = 0.0002$). The machine learning-based analysis

(**Figure 6B**) did not highlight any difference between the AUC values of the two systems, at any time resolution, showing for both the technologies AUC on average higher than 0.8.

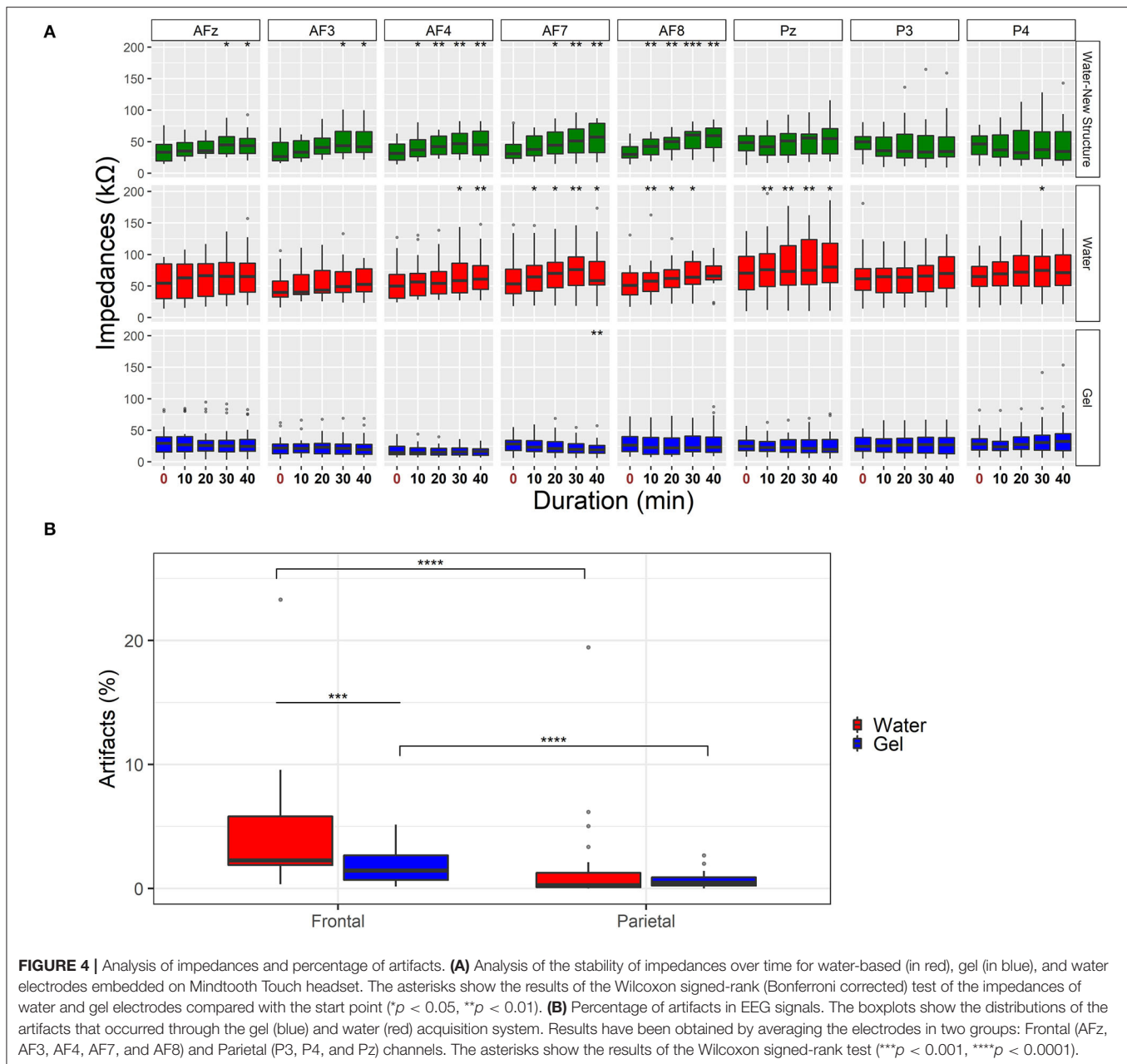
Neurometrics of stress computed by using both technologies showed a significant difference between the low stress and the high-stress conditions (Gel: $Z = 45$, $p = 0.024$; Water: $Z = 3$, $p < 10^{-5}$) (**Figure 7A**). The machine learning-based analysis (**Figure 7B**) highlighted a statistically significant difference between the AUC values of the two systems, starting from 4 s resolution. In particular, the water-based system provided significantly higher AUC than gel-based ones.

Neurometrics of vigilance (**Figure 8A**) computed by using both the two technologies showed a significant difference between the High vigilance and the Low vigilance conditions (Gel: $Z = 179$, $p = 0.0042$; Water: $Z = 183$, $p = 0.0023$). The machine learning-based analysis (**Figure 8B**) did not highlight any difference between the AUC values of the two systems, at any time resolution. The AUC overcame 0.8 at 10 s of temporal resolution.

Study 2: Test on Realistic Driving

Neurometrics of workload computed by using both the two technologies showed a significant difference between the easy and the hard condition (Gel: $Z = 34$, $p = 0.0064$; Water: $Z = 35$, $p = 0.0073$) (**Figure 9A**). The machine learning-based analysis (**Figure 9B**) did not highlight any difference between the AUC values of the two systems, at any time resolution. For both systems, the AUC overcomes 0.8 at 30 s of temporal resolution.

To investigate the ability to generalize the proposed water-based passive-BCI system, the data recorded during the laboratory tasks have been used to calibrate the system. Therefore, it has been tested on simulated driving. Results in **Figure 10A** showed that the system reached an AUC value

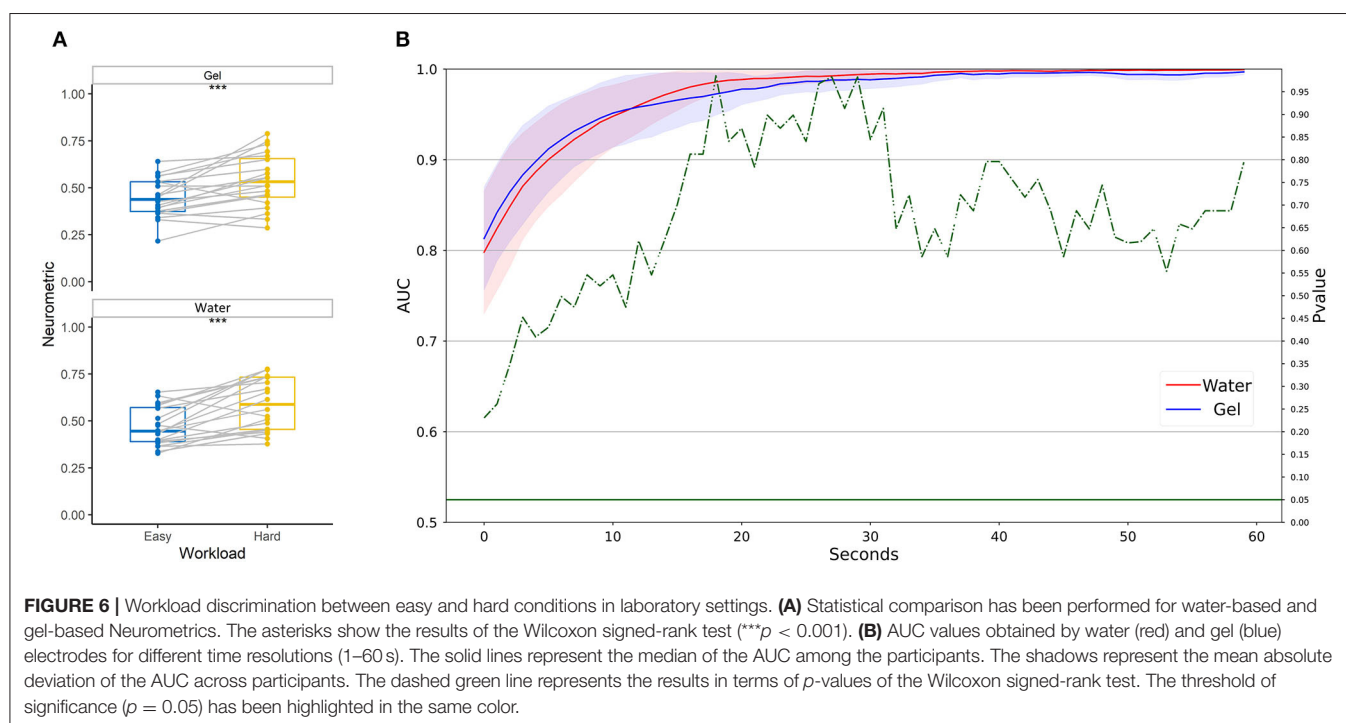
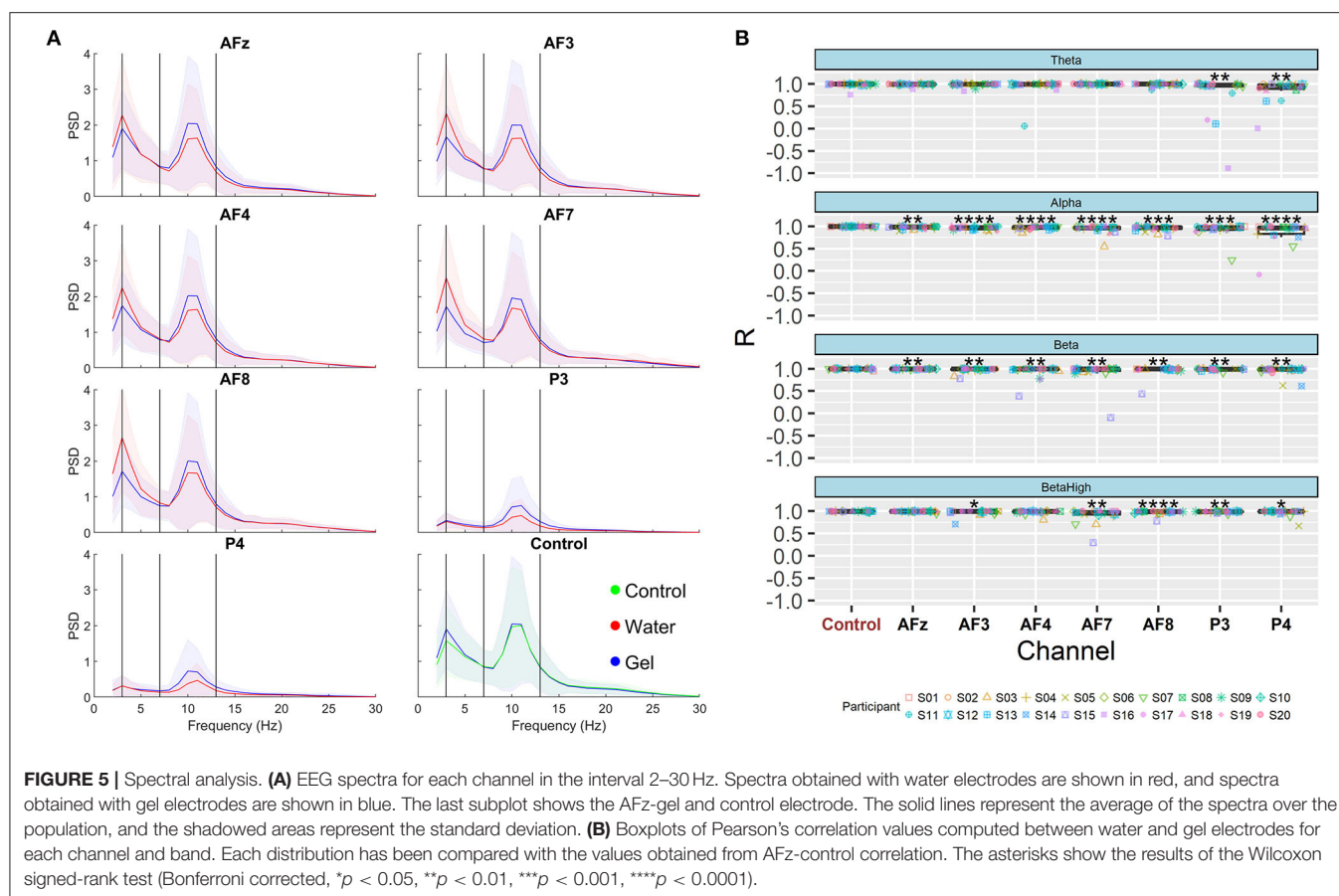


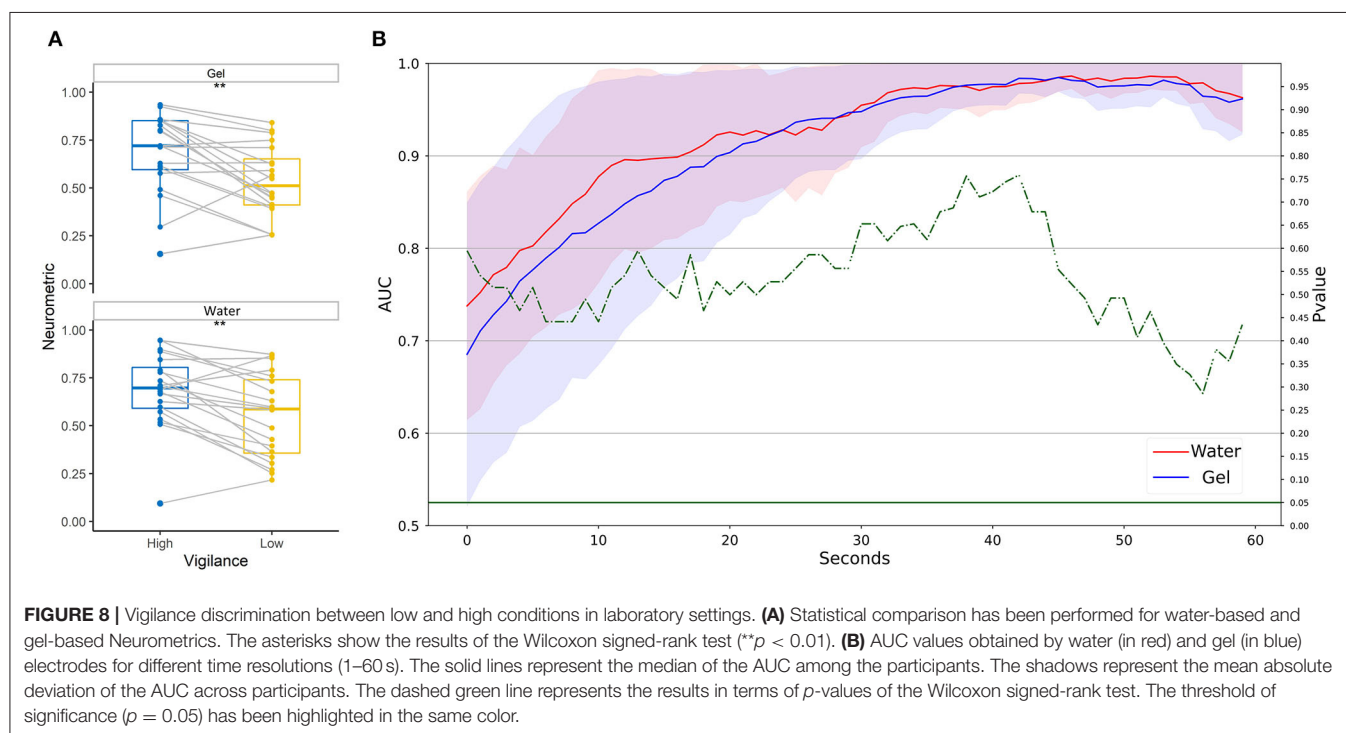
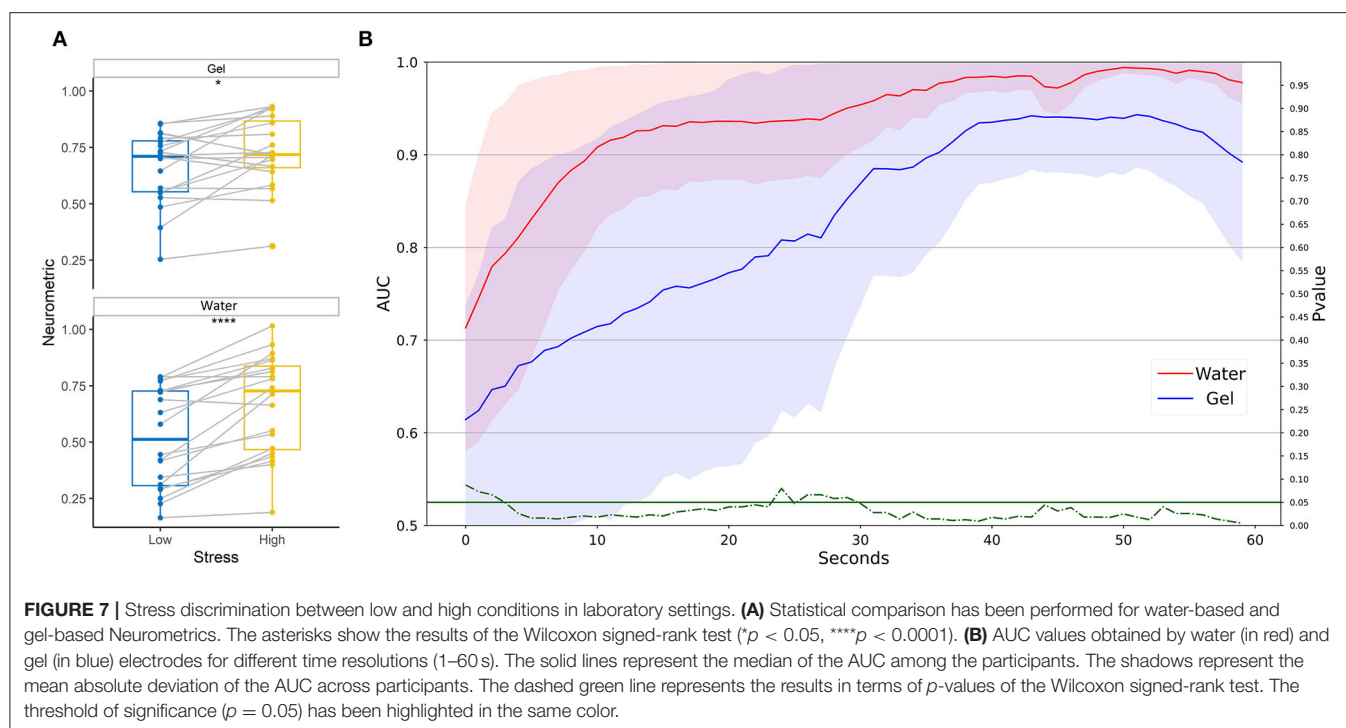
higher than 0.8 with 40 s of temporal resolution. The cross-task approach did not perform significantly differently compared to the intra-subject approach. In **Figure 10B** the related scores obtained at 40 s of temporal resolution indicate that there is a medium correlation between the outputs of the two models. The median over the population of the correlation between the scores obtained is $R = 0.60$ (IQR = 0.54, all $p < 10^{-5}$), and the RMSE = 0.16 (IQR = 0.05).

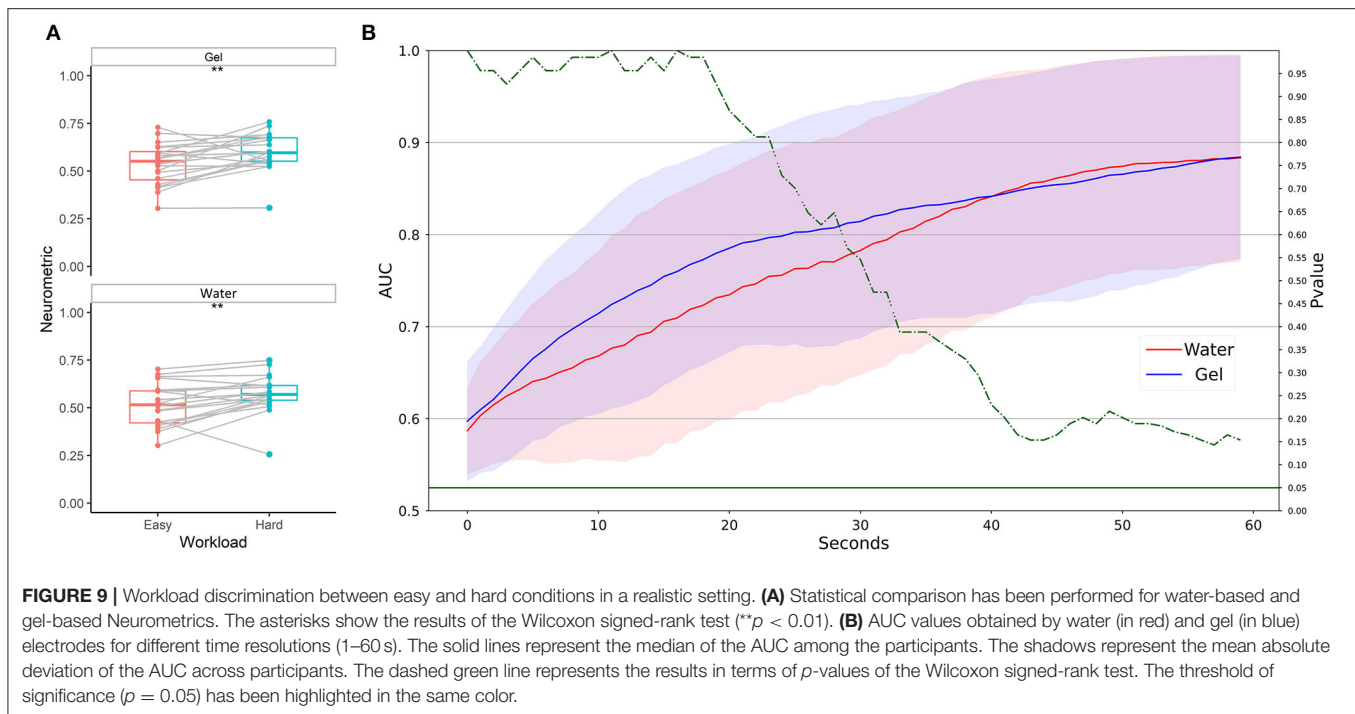
Neurometrics of stress (**Figure 11A**) computed by using both the two technologies showed a significant difference between the low stress and the high-stress conditions (Gel: $Z = 0$, $p < 10^{-5}$; Water: $Z = 0$, $p < 10^{-5}$). The machine learning-based analysis (**Figure 11B**) did not highlight any difference between the AUC

values of the two systems, at any time resolution, showing for both the technologies AUC higher than 0.8 on average.

Even for the stress, it has been investigated the ability to generalize of the proposed water-based passive-BCI system. First, the data recorded during the laboratory multitasking have been used to calibrate the system. Then, the calibrated system has been tested on the simulated driving. Results in **Figure 12A** showed that the system reached AUC values higher than 0.8 almost in real time. The intra-subject calibration performed significantly better than the cross-task approach until a temporal resolution of 45 s. The related scores shown in **Figure 12B** indicated that there is a high correlation between the outputs of the two models. The median (IQR) over the population of the correlation between the







scores obtained is $R = 0.89$ (IQR = 0.28, all $p < 10^{-3}$), and the RMSE = 0.27 (0.16).

Neurometrics of vigilance (**Figure 13A**) computed by using the two technologies showed a significant difference between the High vigilance and the Low vigilance conditions (Gel: $Z = 198$, $p = 0.00013$; Water: $Z = 177$, $p = 0.0056$). The machine learning-based analysis (**Figure 13B**) did not highlight any difference between the AUC values of the two systems, at any time resolution. AUC values higher than 0.8 were reached at 30 s of temporal resolution.

The data recorded during the PVT has been used to calibrate the water-based passive-BCI system. Then, the calibrated system has been tested on the simulated driving. Results in **Figure 14A** showed that the cross-task and the intra-subject calibration did not provide significantly different performances, even if the intra-subject approach provided the highest values of AUC overcoming 0.9. The scores shown in **Figure 14B** have been analyzed in terms of correlation and distances between the curves. The median over the population of the correlation between the scores obtained is $R = 0.48$ (IQR = 0.36, all $p < 0.01$), and the RMSE = 0.19 (0.12).

DISCUSSION

Passive-BCI systems are an effective tool for strengthening human-machine interaction. The use of such technologies has been proved to be particularly useful in those human-centered areas where safety and adaptive training are relevant concerns. However, their employment away from laboratories or research activities has not yet taken place. The reason for this lies in shortcomings in three fundamental aspects: reliability, generalizability, and ease of use. In this context, this work aimed

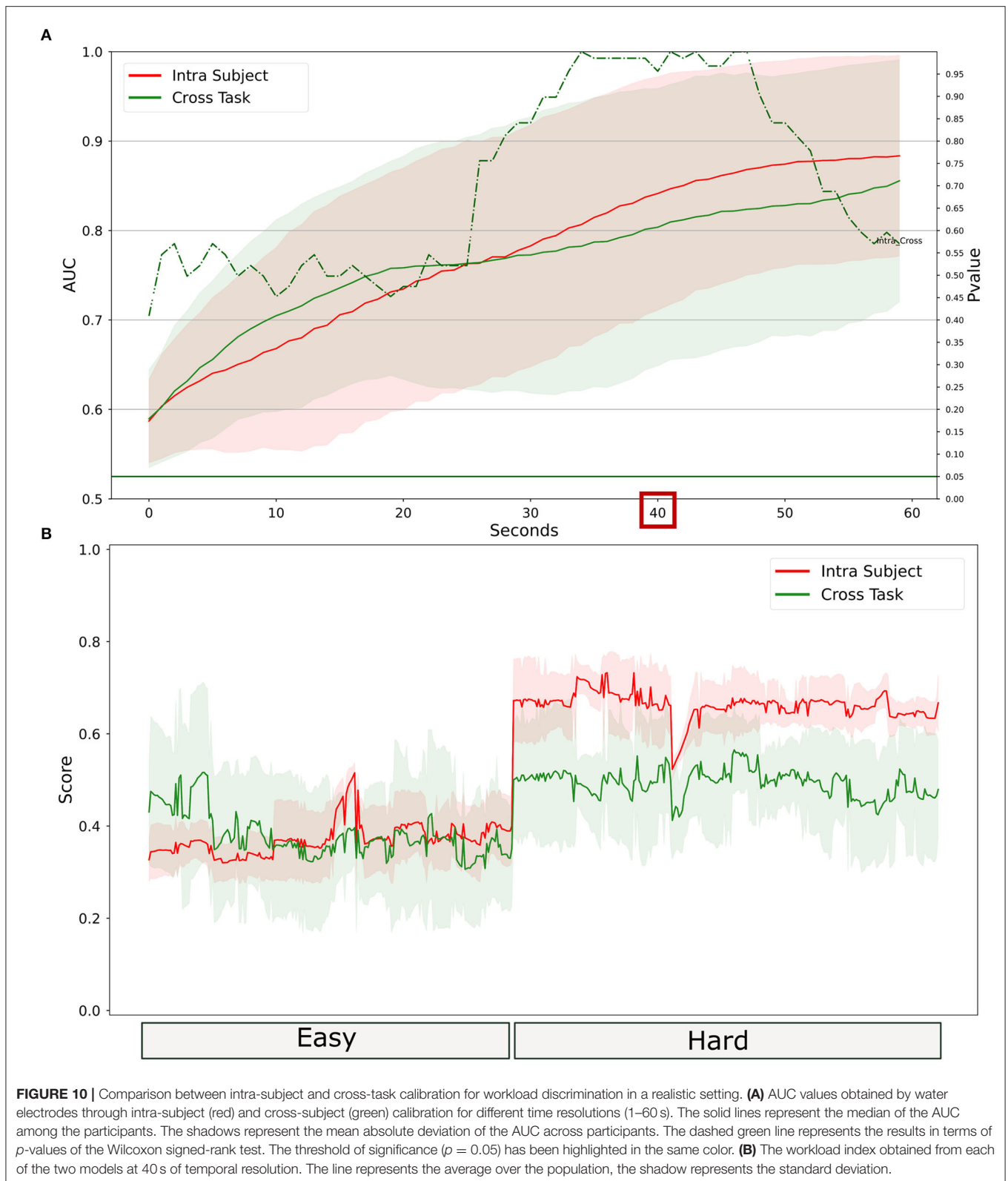
to evaluate a new water-based passive-BCI system covering (i) each of the aforementioned features; (ii) laboratory and realistic settings; and (iii) three different human factors.

Water-Based vs. Gel-Based System: Impact on the Ease of Use and Reliability

The results obtained showed that the use of the new water-based system represents a decisive improvement in terms of usability and that the signal quality guarantees result comparable to those obtained with the gel system.

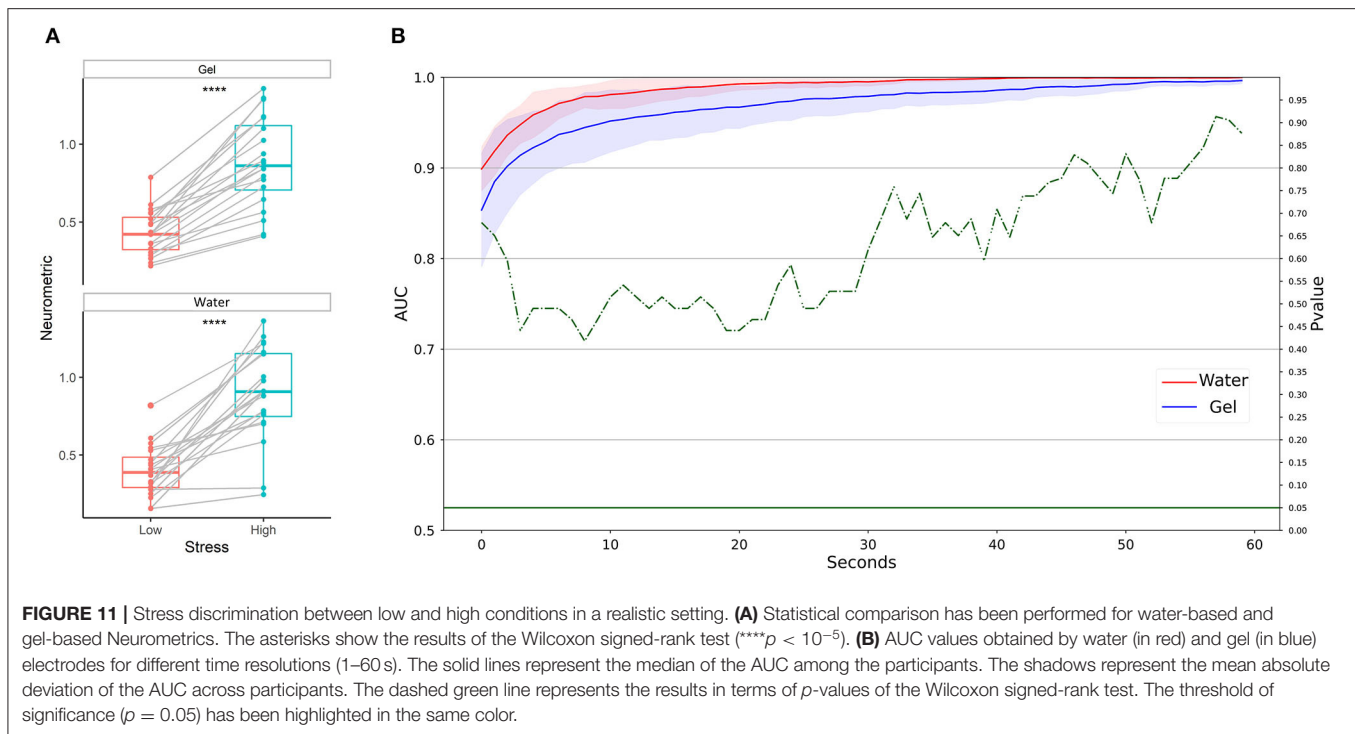
The usability test of the Mindtooth Touch headset embedded with water-based electrodes showed that it allows a quick and easy application, 10 s on average, avoiding every kind of damaging skin preparation. However, one of the participants found it “Hard” to take off the headset (**Figure 3**). This issue was mainly due to the imperfections at the corners of the electrode holders, which could get stuck in the hair. This possible issue will be mitigated by smoothing and rounding edges. No participants found it uncomfortable to wear the headset. After 10 min, 30% considered wearing the headset acceptable. By interviewing the participants, we understood that the comfort issue was mainly due to the pressure induced by the frontal holders. One solution implemented to mitigate this issue was to place a soft foam pad attached with Velcro to the front to evenly distribute pressure over the front head and improve overall comfort. In general, due to the ease of use and comfort, the system proved to be ready for recording EEG out of the lab, and its use will be validated in real settings like pilots and driver training and during working activities.

To analyze the stability of each system during the recording, the impedances of the analyzed systems were evaluated at the



beginning and the end of each experimental condition. The results demonstrated that the impedances of gel electrodes did not significantly vary over time, except for AF7 showed a

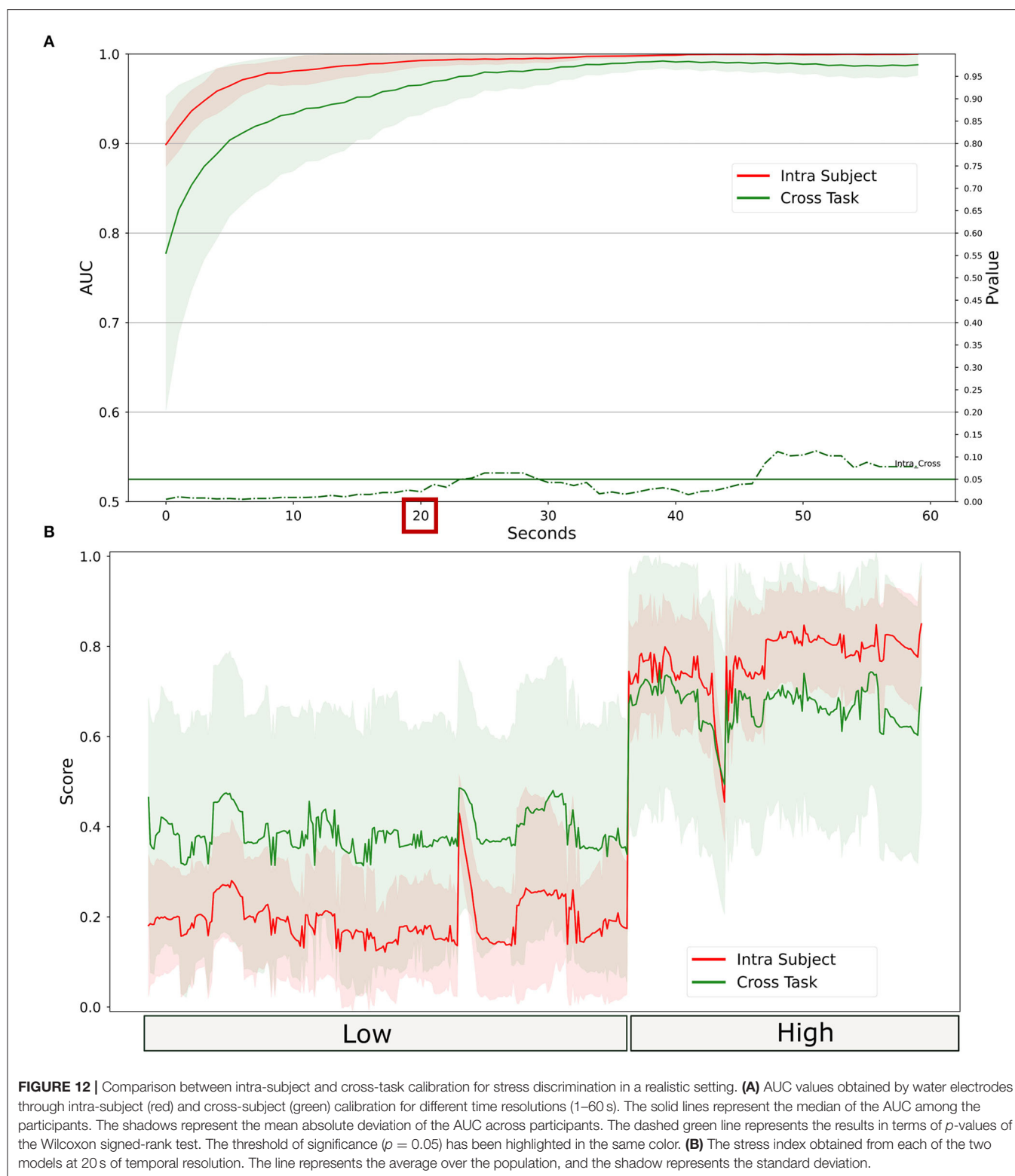
significant reduction compared to the start of the experiment (**Figure 4A**). This is because the impedances of gel electrodes are usually subjected to a natural adaptation before the gel gets dry



(Toyama et al., 2012). The impedances of water electrodes are on average lower than 100 k Ω . Unlike the parietal electrodes, the frontal electrodes showed a significant increment of impedances in the 40-minute window that has been analyzed. This indicates that the electrodes on the forehead were subjected to a faster drying than the parietal, whereas the presence of hair can prevent this from happening. However, these differences did not depend only on the electrochemical characteristics of the electrodes, but also on the mechanical fixation. Analyzing the same water electrodes placed on the specially designed structure, the impedances decreased to 50 k Ω , even if they showed the same increase on the frontal sites over time. As seen so far, the Mindtooth Touch structure in purpose has been specifically designed to facilitate the use of water electrodes while promoting comfort, time spent fitting the headset, and ease of use. However, as the recording advances, and the saline solution evaporates, impedances increase over time and this can negatively affect the data quality of long recordings (Noreika et al., 2020). For long recording (more than 30 min) the saline solution could be reapplied using a pipette and without removing the headset.

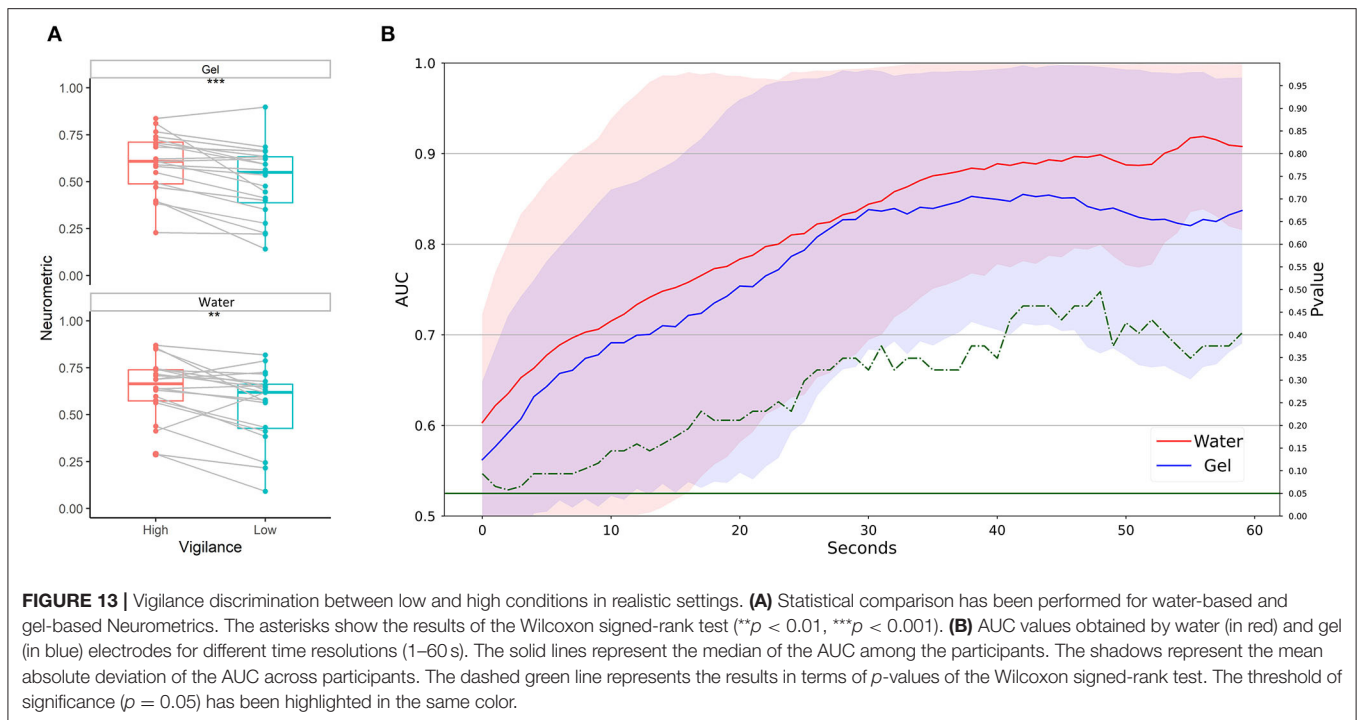
Higher impedances characterizing the water-based system could impact on signal to noise ratio (SNR) of EEG signals because of the common-mode rejection and cephalic skin potentials. Common mode rejection refers to the ability of a recording system to reject noise that is in common with the recording sensors and reference electrodes. As the electrode impedance increases, the common-mode rejection of the system decreases, and the SNR of the recording decreases as well. To deal with the problem of decreased common-mode rejection with high electrode impedances, amplifiers with a

higher input impedance have been used. In this regard, the used LiveAmp amplifier has a very high input impedance (i.e., 200 M Ω), so this first problem could be automatically solved. Electrode impedances that are too high may lead to a second problem that cannot be solved using changes to the amplifier's input impedance, namely, an increase in skin potential artifacts. Skin potentials arise because of the standing electrical potential that is normally present between the inside and the outside of the skin (Edelberg, 1972). Current literature demonstrated that high impedance values (i.e., 200 k Ω) could increase noise and induce changes in a frequency lower than 5 Hz since skin potentials induce slow voltage variations (Ferree et al., 2001; Kappenman and Luck, 2010). In this regard, both systems showed a higher number of artifacts on frontal electrodes compared to parietal ones (Figure 4B). Frontal electrodes seem more sensitive to unrelated electrical signals originating from non-brain physiological activities in particular muscular and ocular activities, reflecting a non-homogenous distribution of artifacts over the scalp (Abdi-Sargezeh et al., 2021). Even if both systems showed a percentage of artifacts on average that is lower than 5% of the total amount of data recorded, the frontal water electrodes recorded a significantly higher number of artifacts compared to gel ones. In contrast, the number of artifacts on parietal sites is not different between the two systems. The higher number of artifacts on frontal sites could be due to the mechanical fixation of the used EEG cap that disadvantaged the water electrodes. As demonstrated, water sensors placed on the Mindtooth Touch structure guarantee better contact and lower impedances.



From the spectral perspective, we investigated if the higher impedance values of the water-based technology compared to the gel-based technology would negatively affect the spectral

features used to compute EEG neurometrics. Visual inspection of the spectra obtained through water and gel electrodes showed that the shape is very similar (**Figure 5A**). The most evident



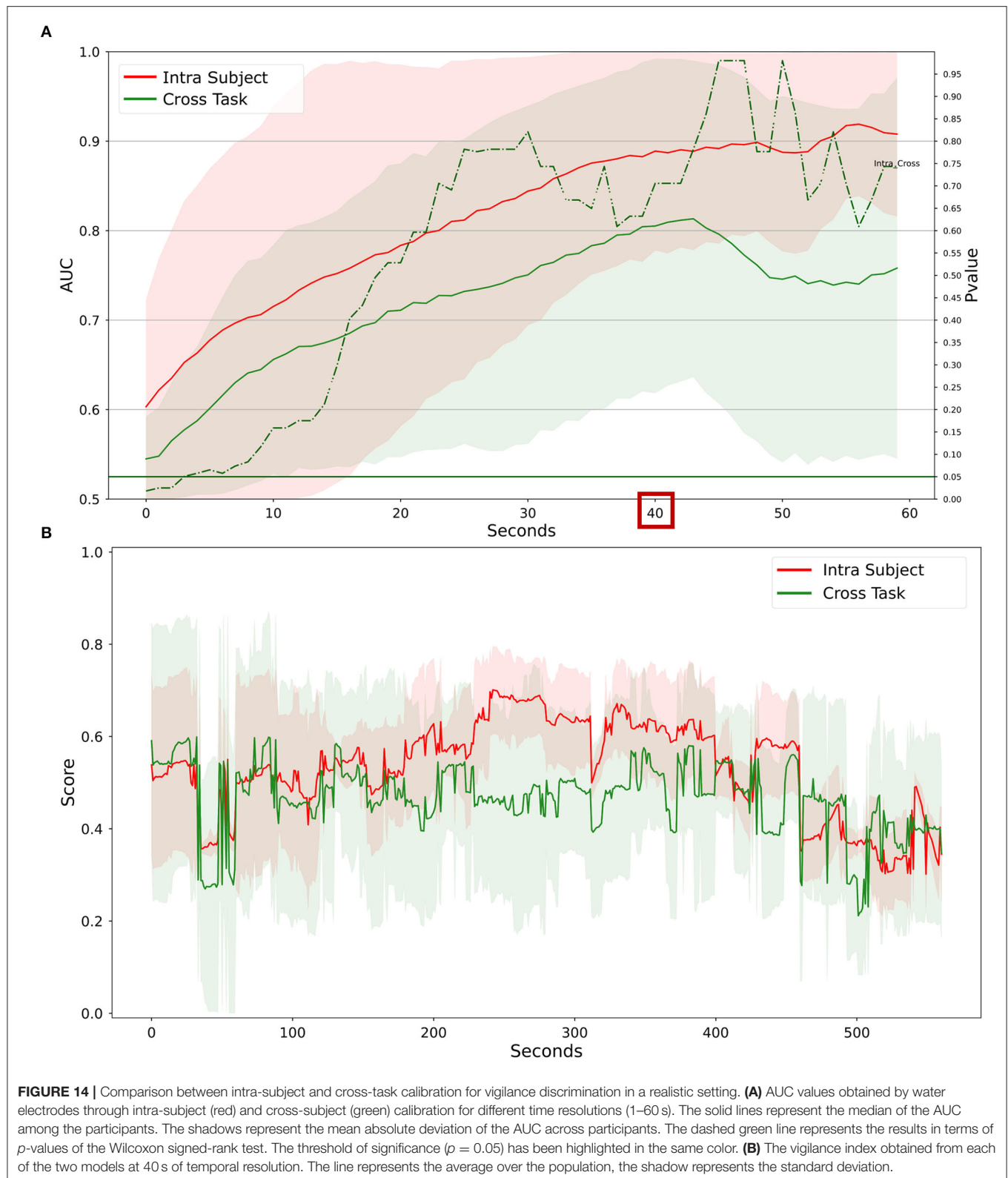
difference is that water electrodes showed higher values in the theta band, while gel showed higher values in the alpha band. The correlation analysis of specific EEG features (e.g., theta, alpha, beta, and high beta bands) during open eyes conditions suggested that the correlation between gel and water features is high ($R > 0.95$). Comparing the distribution of R values with that obtained from correlating the AFz-gel electrode with the Control electrode (correlation between two gel electrodes was used as a benchmark, therefore eventually differences in correlation depend on the distances between the two electrodes), we found that even if the correlation between gel and water spectral features is high, it is significantly lower compared to the control-AFz spectra especially for alpha and beta bands, whereas this is less evident in the theta band. These differences found between the different bands are coherent with similar spectral comparisons performed between a gel-based and water-based system (Topor et al., 2021). The authors found that the water-based system behaves significantly different from the gel-based in the detection of high-beta desynchronization, that the theta band was the less consistent between the two systems, and that there was a frontal shift of maximal power in theta and alpha bands. A slow drift noise has been hypothesized to be the cause of this. This correlation analysis has been performed on a highly controlled condition (i.e., open eyes), however, the degree of freedom during the analyzed paradigms could impact the spectral correlation depending on the type of electrode (Tautan et al., 2014). The possible loss of quality of EEG signal recorded from water-based electrodes could not impact neurometrics assessment when the experimental task is very controlled (i.e., the participant does not move too much, simple tasks, low number of expected artifacts), but it could affect the computation in real settings (i.e.,

higher number of expected artifacts, movement of the head, and real tasks). For this reason, the employment of the proposed water-based passive BCI has been tested in both laboratory and realistic settings.

Study 1: Human Factors Discrimination During Laboratory Tasks

The aim of Study 1 is to test the proposed system during controlled laboratory tasks and to use these data to calibrate the passive-BCI system based on a machine learning model. Before going on to test a system during a realistic application, the step to a more controlled application is a must in the case of passive-BCI. One of the reasons is that the targeted human factors are not independent variables. For example, stress is affected by workload level, conversely, the effort involved in coping with stress actually adds to the task demands (Stanton and Young, 2000); a variation in workload is associated with a variation in fatigue (Roy et al., 2016), and so on. Thanks to controlled tasks performed in the laboratory it is more likely to induce a variation of a single human factor per time compared to realistic contexts. Therefore, two standard tasks have been used to elicit workload, stress, and vigilance variations.

Multitasking has been chosen as a laboratory task to induce workload and stress modulation. This task is similar to the simultaneous capacity (SIMKAP) task usually used to elicit workload variation and whose effectiveness in eliciting stress variation has been already proved (Wetherell and Sidgreaves, 2005; Lim et al., 2018). Multitasking has been preferred to operate-orientated task paradigms like the multi-attribute task battery (MATB) because it consists of common tasks like entry numbers, mathematical additions, bar filling, and



auditory test, and the time of familiarization is shorter than those required for the MATB (Comstock, 1992). Compared to multitasking, the MATB was born in the context of

aviation research and usually requires 1 week of training to avoid any learning bias. As the employed task should be used to calibrate a cross-task system to be used in a

realistic environment, it is necessary that the training is fast and straightforward.

The results obtained on multitasking showed that the neurometrics obtained with both gel and water electrodes were able to significantly discriminate the high level of the workload from the low. This confirms previous research, where a measure of workload based on frontal activity in the theta band and parietal activity in the alpha band has been consolidated and applied in different contexts (Borghini et al., 2014). The same neurophysiological information used as features to calibrate a machine learning model based on a random forest showed AUC values higher than 0.8 for high temporal resolution, and values of almost 1 for resolution lower than 30 s in the laboratory setting. There is not a statistically significant difference between the performance obtained with the two systems. On a similar multitasking dataset, 69.2% of accuracy has been reached through a support vector machine (Lim et al., 2018) and 82.57% accuracy using the LSTM model (Chakladar et al., 2020).

Analogously, the neurometrics obtained with gel and water electrodes were both able to significantly discriminate between high and low stress during the multitasking levels. Moreover, the discrimination of stress levels during the laboratory task through the machine learning-based approach showed significantly different performances between the water and the gel systems. In particular, the AUC values obtained with the water-based system are significantly higher than those obtained with the gel ones by about 10%. The proposed passive BCI system employed to discriminate between two levels of stress showed higher performance compared with the gold standard during the laboratory experiment. We could hypothesize that this increment could be associated with spectral differences discovered during spectral analysis. Comparing these outcomes with the literature, Attallah (2020) reached an accuracy of 99.26% for discriminating between two levels of stress with a KNN model.

Finally, both systems provided neurometrics that can significantly discriminate the high level of vigilance from the low during the PVT. The discrimination of the two levels of vigilance during the PVT, through the machine learning-based approach provided AUC values higher than 0.7 on average for high temporal resolution, and values higher than 0.95 for resolution lower than 30 s in the laboratory setting. There is not a statistically significant difference between the performance obtained with the two systems. Therefore, the proposed passive BCI system employed to discriminate between two levels of vigilance did not behave differently from the gold standard. The results obtained are in line with the literature: using an SVM model, Mehreen et al. (2019) obtained 76% of accuracy, while Wei et al. (2018) reached 0.85 of AUC using a non-hair bearing EEG. In Choi et al. (2019), the PVT task has been used to effectively train a system to detect instantaneous drowsiness with EEG. Using an extreme gradient boosting as a machine learning classifier, the authors tested both a wired/wet EEG and a wireless/dry EEG verifying that the latter reached 0.81 of AUC, 7 % less than the standard system.

In conclusion, analyzing the results of Study 1 the proposed system did not behave differently from the gel-based system when neurometrics are used to discriminate between human factors levels. Moreover, water-based electrodes overcame the

traditional ones for the discrimination of the two levels of stress and did not behave significantly differently in case of vigilance or workload discrimination.

Study 2: Human Factors Discrimination During Realistic Driving

The neurometrics obtained with both gel and water electrodes were able to significantly discriminate the high level of the workload from the low during realistic driving. Analogously, the workload during the driving task has been discriminated with AUC around 0.6 for high resolution (i.e., 1 s) and the threshold of 0.8 was surpassed only after 30 s of time resolution. Also, in this case, there was no significant difference between the two systems. Therefore, the proposed passive BCI system employed to discriminate between two levels of workload showed higher performance in laboratory settings compared to the realistic ones for the same time resolution. However, it did not behave differently from the gold standard. Comparing the obtained results with the literature, 72% accuracy has been obtained through a support vector machine and a driving simulator (Hernández et al., 2018). The author (Fan et al., 2015) obtained an accuracy of 81.46% using a KNN and a driving system in virtual reality.

The neurometrics obtained with gel and water electrodes were both able to significantly discriminate between high and low stress in realistic settings. Moreover, there is not a statistically significant difference between the performance obtained with the two machine learning based systems during the realistic driving. Both systems showed an AUC higher than 0.9 for a temporal resolution that is lower than 4 s. Therefore, the proposed passive BCI system employed to discriminate between two levels of stress, showed higher performance compared with the gold standard during the laboratory experiment, while they behave similarly during realistic tasks. We could hypothesize that this minimal decrease could be associated with the differences in the position of the electrodes, thus, representing one limitation of the current study. In Halim and Rehan (2020) more than 50 automotive drivers have been tested during various driving situations. An SVM model reached 97.95% of accuracy in discriminating rest and stress during driving through EEG signals.

The neurometrics obtained with gel and water electrodes were both able to significantly discriminate the high level of vigilance from the low ones in realistic settings. The vigilance during the driving task has been discriminated with AUC around 0.6 for high temporal resolution (i.e., 1 s) and the water-based electrodes allowed to overcome 0.9 on average AUC for temporal resolution lower than 50 s. Therefore, the proposed passive BCI system that was employed to discriminate between two levels of vigilance showed higher performance in a laboratory setting compared to the real ones for the same time resolution. However, it did not behave differently from the gold standard. Similar results have been already obtained through dry electrodes in the driving context. A case study involving 15 participants in an immersive virtual driving environment demonstrates the reliability and the feasibility of predicting the driver's vigilance through support vector regression (Lin et al., 2014).

Generalizability of the Cross-Task Approach

The generalizability of a passive-BCI system depends strictly on the implemented algorithms, in terms of features and models. Generalizability has been defined as the algorithm's ability to generalize across tasks, sessions, or subjects with an accuracy higher than 80% (Zhou et al., 2021). In this work, we focused on cross-task generalizability. For all the targeted human factors, the water-based passive-BCI system has been calibrated using laboratory data. The design of appropriate training tasks for a cross-task approach is the main challenge since the tasks should be simple, short, and without confounding effects (Gerjets et al., 2014).

The obtained results showed that the water-based passive-BCI system calibrated cross-task allowed the discrimination between two levels of the targeted human factors. In particular, this approach applied to the workload monitoring reached an AUC higher than 0.8 at 40 s of temporal resolution providing the same performance as the intra-subject approach. Similar results have been obtained for the monitoring of vigilance. The intra-subject approach provided performances 10% higher than the cross-task approach, even if they are not significantly different. The highest values in terms of AUC have been reached during the stress monitoring: in both cases, the systems reached values higher than 0.95 after 20 s of temporal resolution. As said, these results are affected by the choice of the task for calibration, therefore the results indicated that the choice made for stress is as far better than those made for workload and vigilance.

The approach used here allowed a continuous estimate of the targeted human factors in terms of score (i.e., the probability of belonging to the high class), which can be considered the output index of the passive-BCI system. We have analyzed the similarity between the indices obtained through the intra-subject and the cross-subject approach. In particular, the stress indices were highly correlated to the population ($R = 0.89$), whereas they showed also higher distances ($RMSE = 0.27$) between the curves compared to the indices obtained for vigilance and workload. In contrast, the correlation between the workload and vigilance indices obtained with the intra-subject and cross-subject approach is medium (0.6 and 0.48, respectively). The temporal delay between the training and testing data may have disadvantaged the cross-task results. Future testing should involve cross-session recordings to better investigate this bias. Overall, the obtained results confirmed the appropriateness of the cross-task approach used for stress assessment, and that improvement should be achieved for workload and vigilance. In future work, different calibration tasks will be designed and tested on realistic driving.

Contextualizing these results in the literature, acceptable values of performance have been found only for cross-task calibrated and then tested on similar tasks. For example, Ke et al. (2015) reached an accuracy higher than 0.9 in mental workload classification using spatial and verbal n-back. In contrast, the study of Baldwin and Penaranda (2012) showed low accuracy in the cross-task classification using three different working memory tasks. The authors hypothesized that this result was due to the different neural structures underpinning each

task. Therefore, when the cross-task procedure encompasses a predefined set of features, on the one hand, this can avoid the misclassification due to diverse concurrent mental states. On the other hand, a predefined set of features limits the generalizability of the cross-task approach. Therefore, in future attempts, the two approaches (predefined set of feature vs. features selection) should be compared.

Limitations and Future Work

One limitation of this study is related to the choice of performing simultaneous recording with the two kinds of electrodes. On the one hand, the simultaneous registration allows for avoiding temporal bias. On the other hand, there is a spatial bias that can affect the obtained correlation because they record different bioelectrical activities. The distance between the electrodes was about 1.5 cm to avoid electrical bridges due to the simultaneous presence of gel and water. A second limitation is related to the specificity of neurophysiological indices in realistic contexts. For example, visual-related activity during the task could hide a variation of parietal alpha due to workload. Also, systematic artifacts could be erroneously selected as features by the system leading to incorrect classification. Moreover, the current paper showed the system can discriminate between two levels of workload, stress, or vigilance. However, it is essential to move from a binary classification to multi-class classification. For example, the well-known inverted-U function of the Yerkes–Dodson law (Yerkes and Dodson, 1908) associated arousal with performance. According to this, it would be ideal to distinguish at least a suboptimal, an optimal, and an overload condition.

In the future, guidelines should be provided for the description of the tasks to be performed for effective cross-task calibration. The task in question must be simple to not require too much time for training, although it must be long enough to guarantee enough data for training the model. Finally, these water-based electrodes should be tested against already existing water electrodes (Volosyak et al., 2010; Schwarz et al., 2020).

CONCLUSION

The main aim of this work was to test a new water-based passive BCI system for workload, vigilance, and stress monitoring for out-of-the-lab applications. The results showed that water electrodes guarantee higher ease of use, without lowering the performance compared to the gold standard. In addition, a proper structure for housing the EEG equipment would make the system wearable, so easy to fit and comfortable, and consequently acceptable for many different real-life applications. Moreover, the generalizability of the system, analyzed through a cross-task approach, showed acceptable performance according to different values of temporal resolution for each investigated human factor.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because of privacy restrictions. Requests to access the datasets should be directed to PA, pietro.arico@uniroma1.it.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board of Sapienza University of Rome. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Conceptualization and writing—original draft preparation: NS and PA. Methodology: NS, AD, and PA. Formal analysis: NS, DG, GD, AG, and PA. Software: AD. Investigation: NS, PA, GD, and GB. Resources and funding acquisition: FB. Data curation: GB and AV. Writing—review and editing: NS, PA, VR, AV, GD, DG, GB, and AG. Supervision: FB and PA. Project administration: FB, PA, GB, and GD. All authors contributed to the article and approved the submitted version.

FUNDING

This work was co-financed by the European Commission by Horizon2020 projects MINDTOOTH: wearable device to decode human mind by neurometrics for a new concept of smart interaction with the surrounding environment (GA No. 950998). H2020-SESAR-2019-2 projects: Transparent artificial intelligence and automation to air traffic management systems,

ARTIMATION, (GA No. 894238), WORKINGAGE: smart working environments for all Ages (GA No. 826232), FITDRIVE: monitoring devices for overall FITness of Drivers (GA No. 953432), SAFEMODE: Strengthening synergies between Aviation and maritime in the area of human Factors toward achieving more Efficient and resilient MODE of transportation (GA No. 814961), and BRAIN-SAFEDRIVE: a technology to detect Mental States during Drive for improving the Safety of the road (Italy-Sweden collaboration) with a grant of Ministero dell'Istruzione dell'Università e della Ricerca della Repubblica Italiana. The individual grants BRAIN2GETHER (BE-FOR-ERC) and NEUROSIM (Avvio alla ricerca 2020), recognized by Sapienza University of Rome to GD are also acknowledged.

ACKNOWLEDGMENTS

The authors sincerely acknowledge the full consortium of the SIMUSAFE project (EU Grant nr. 723386) and, in particular, the Instituto Tecnológico de Castilla y León (ITCL), the MOSS Group of the University Gustave Eiffel (UGE) of Paris, and the Computer Science Laboratory of the University of Porto (FEUP), for making available and setting their car simulator for this study. The authors are grateful to BrainProducts GmbH (Germany, EU) for loaning their instrumentation and for allowing the publication of the obtained results.

REFERENCES

- Abdi-Sargezeh, B., Foodeh, R., Shalchyan, V., and Daliri, M. R. (2021). EEG artifact rejection by extracting spatial and spatio-spectral common components. *J. Neurosci. Methods* 358, 109182. doi: 10.1016/j.jneumeth.2021.109182
- Alimardani, M., and Hiraki, K. (2020). Passive brain-computer interfaces for enhanced human-robot. *Interaction* 7, 1–12. doi: 10.3389/frobt.2020.00125
- Amaral, C. P., Simões, M. A., Mouga, 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
- Aricò, P., Borghini, G., Di Flumeri, G., Bonelli, S., Golfetti, A., Graziani, I., et al. (2017b). Human factors and neurophysiological metrics in air traffic control: a critical review. *IEEE Rev. Biomed. Eng.* 10, 250–263. doi: 10.1109/RBME.2017.2694142
- Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Pozzi, S., and Babiloni, F. (2016). A passive brain-computer interface application for the mental workload assessment on professional air traffic controllers during realistic air traffic control tasks. *Prog. Brain Res.* 228, 295–328. doi: 10.1016/bs.pbr.2016.04.021
- Aricò, P., Borghini, G., Di Flumeri, G., Sciaraffa, N., and Babiloni, F. (2018). Passive BCI beyond the lab: current trends and future directions. *Physiol. Meas.* 39, 08TR02. doi: 10.1088/1361-6579/aad57e
- Aricò, P., Borghini, G., Di Flumeri, G., Sciaraffa, N., Colosimo, A., and Babiloni, F. (2017a). Passive BCI in operational environments: insights, recent advances, and future trends. *IEEE Trans. Biomed. Eng.* 64, 1431–1436. doi: 10.1109/TBME.2017.2694856
- Attallah, O. (2020). An effective mental stress state detection and evaluation system using minimum number of frontal brain electrodes. *Diagnostics* 10, 292. doi: 10.3390/diagnostics10050292
- Baldwin, C. L., and Penaranda, B. N. (2012). Adaptive training using an artificial neural network and EEG metrics for within-and cross-task workload classification. *NeuroImage* 59, 48–56. doi: 10.1016/j.neuroimage.2011.07.047
- Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J. Math. Psychol.* 12, 387–415.
- Blankertz, B., Acqualagna, L., Dähne, S., Haufe, S., Schultze-Kraft, M., Sturm, I., et al. (2016). The Berlin brain-computer interface: progress beyond communication and control. *Front. Neurosci.* 10, 530. doi: 10.3389/fnins.2016.00530
- Borghini, G., Aricò, P., Astolfi, L., Toppi, J., Cincotti, F., Mattia, D., et al. (2013). “Frontal EEG theta changes assess the training improvements of novices in flight simulation tasks,” in *Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 6619–6622. doi: 10.1109/EMBC.2013.6611073
- 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
- Borghini, G., Di Flumeri, G., Aricò, P., Sciaraffa, N., Bonelli, S., Ragosta, M., et al. (2020). A multimodal and signals fusion approach for assessing the impact of stressful events on air traffic controllers. *Sci. Rep.* 10, 1–18. doi: 10.1038/s41598-020-65610-z
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324
- Chakladar, D., Das, Dey, S., Roy, P. P., and Dogra, D. P. (2020). EEG-based mental workload estimation using deep BLSTM-LSTM network and evolutionary algorithm. *Biomed. Signal Process. Control* 60, 101989. doi: 10.1016/j.bspc.2020.101989
- Chi, Y. M., Wang, Y., Te, Wang, Y., Maier, C., Jung, T. P., and Cauwenberghs, G. (2012). Dry and noncontact EEG sensors for mobile brain-computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* 20, 228–235. doi: 10.1109/TNSRE.2011.2174652
- Choi, H. S., Min, S., Kim, S., Bae, H., Yoon, J. E., Hwang, I., et al. (2019). Learning-based instantaneous drowsiness detection using wired and wireless electroencephalography. *IEEE Access* 7, 146390–146402. doi: 10.1109/ACCESS.2019.2946053

- Comstock, Jr., Raymond, J., and Arnegard, R. J. (1992). *The Multi-Attribute Task Battery for Human Operator Workload and Strategic Behavior Research*. No. NAS 1.15: 104174.
- 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., Colosimo, A., and Babiloni, F. (2016). “A new regression-based method for the eye blinks artifacts correction in the EEG signal, without using any EOG channel,” in *Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society Conference*. doi: 10.1109/EMBC.2016.7591406
- Di Flumeri, G., Aricò, P., Borghini, G., Sciaraffa, N., Di Florio, A., and Babiloni, F. (2019a). 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
- Di Flumeri, G., Borghini, G., Aricò, P., Sciaraffa, N., Lanzi, P., Pozzi, S., et al. (2018). EEG-based mental workload neurometric to evaluate the impact of different traffic and road conditions in real driving settings. *Front. Hum. Neurosci.* 12, 509. doi: 10.3389/fnhum.2018.00509
- Di Flumeri, G., De Crescenzo, F., Berberian, B., Ohneiser, O., Kramer, J., Aricò, P., et al. (2019b). Brain–computer interface-based adaptive automation to prevent out-of-the-loop phenomenon in air traffic controllers dealing with highly automated systems. *Front. Hum. Neurosci.* 13, 296. doi: 10.3389/fnhum.2019.00296
- Di Nardo, M. D., Forino, D., and Murino, T. (2020). The evolution of man–machine interaction: the role of human in Industry 4.0 paradigm. *Prod. Manuf. Res.* 8, 20–34. doi: 10.1080/21693277.2020.1737592
- Douibi, K., Le Bars, S., Lemontey, A., Nag, L., Balp, R., and Breda, G. (2021). Toward EEG-based BCI applications for industry 4.0: challenges and possible applications. *Front. Hum. Neurosci.* 15, 705064. doi: 10.3389/fnhum.2021.705064
- Edelberg, R. (1972). Electrical activity of the skin: its measurement and uses in psychophysiology. *Handbook Psychophysiol.* 1972:367–418.
- Erp, J. B. F., Van, Lotte, F., and Tangermann, M. (2012). Brain–computer interfaces: Beyond medical applications. *Computer*. 45, 26–34. doi: 10.1109/MC.2012.107
- Fan, J., Wade, J. W., Bian, D., Key, A. P., Warren, Z. E., Mion, L. C., et al. (2015). “A step towards EEG-based brain computer interface for autism intervention,” in *Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 3767–3770.
- Ferree, T. C., Luu, P., Russell, G. S., and Tucker, D. M. (2001). Scalp electrode impedance, infection risk, and EEG data quality. *Clin. Neurophysiol.* 112, 536–544. doi: 10.1016/S1388-2457(00)00533-2
- Gerjets, P., Walter, C., Rosenstiel, W., Bogdan, M., and Zander, T. O. (2014). Cognitive state monitoring and the design of adaptive instruction in digital environments: lessons learned from cognitive workload assessment using a passive brain-computer interface approach. *Front. Neurosci.* 8, 385. doi: 10.3389/fnins.2014.00385
- Gevins, A., Smith, M. E., McEvoy, L., and Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. *Cerebral Cortex* 7, 374–385.
- Halim, Z., and Rehan, M. (2020). On identification of driving-induced stress using electroencephalogram signals: A framework based on wearable safety-critical scheme and machine learning. *Information Fusion* 53, 66–79. doi: 10.1016/j.inffus.2019.06.006
- He, H., Bai, Y., García, E. A., and Li, S. (2008). ADASYN: adaptive synthetic sampling approach for imbalanced learning. *Proc. Int. Joint Conf. Neural Netw.* 3, 1322–1328. doi: 10.1109/IJCNN.2008.4633969
- Hekmatmanesh, A., Zhidchenko, V., Kauranen, K., Siitonen, K., Handroos, H., Soutukorva, S., et al. (2021). Biosignals in human factors research for heavy equipment operators: a review of available methods and their feasibility in laboratory and ambulatory studies. *IEEE Access* 9, 97466–97482. doi: 10.1109/ACCESS.2021.3092516
- Hernández, L. G., Mozos, O. M., Ferrández, J. M., and Antelis, J. M. (2018). EEG-based detection of braking intention under different car driving conditions. *Front. Neuroinform.* 12, 1–14. doi: 10.3389/fninf.2018.00029
- Hobfoll, S. E., and Shirom, A. (1993). Stress and burnout in the workplace: conservation of resources. *Handbook Organ. Behav.* 1, 41–61.
- Hubbard, J., Kikumoto, A., and Mayr, U. (2019). EEG decoding reveals the strength and temporal dynamics of goal-relevant representations. *Sci. Rep.* 9, 1–11. doi: 10.1038/s41598-019-45333-6
- Jamil, N., Belkacem, A. N., Ouhbi, S., and Lakas, A. (2021). Noninvasive electroencephalography equipment for assistive, adaptive, and rehabilitative brain–computer interfaces: a systematic literature review. *Sensors* 21, 4754. doi: 10.3390/s21144754
- Kamrud, A., Borghetti, B., Kabban, C. S., and Miller, M. (2021). Generalized deep learning EEG models for cross-participant and cross-task detection of the vigilance decrement in sustained attention tasks. *Sensors*. 21, 5617. doi: 10.3390/s21165617
- Kappenman, E. S., and Luck, S. J. (2010). The effects of electrode impedance on data quality and statistical significance in ERP recordings. *Psychophysiology* 47, 888–904. doi: 10.1111/j.1469-8986.2010.01009.x
- Katmah, R., Al-Shargie, F., Tariq, U., Babiloni, F., Al-Mughairbi, F., and Al-Nashash, H. (2021). A review on mental stress assessment methods using eeg signals. *Sensors* 21, 5043. doi: 10.3390/s21155043
- Ke, J., Zhang, M., Luo, X., and Chen, J. (2021). Monitoring distraction of construction workers caused by noise using a wearable electroencephalography (EEG) device. *Autom. Const.* 125, 103598. doi: 10.1016/j.autcon.2021.103598
- Ke, Y., Qi, H., Zhang, L., Chen, S., Jiao, X., Zhou, P., et al. (2015). Towards an effective cross-task mental workload recognition model using electroencephalography based on feature selection and support vector machine regression. *Int. J. Psychophysiol.* 98, 157–166. doi: 10.1016/j.ijpsycho.2015.10.004
- Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res. Brain Res. Rev.* 29, 169–195.
- Li, G., and Chung, W.-Y. (2022). Electroencephalogram-based approaches for driver drowsiness detection and management: a review. *Sensors* 22:1100. doi: 10.3390/s22031100
- Lim, W. L., Sourina, O., and Wang, L. P. (2018). STEW: simultaneous task EEG workload data set. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 2106–2114. doi: 10.1109/TNSRE.2018.2872924
- Lin, C. T., Chuang, C. H., Huang, C. S., Tsai, S. F., Lu, S. W., Chen, Y. H., et al. (2014). Wireless and wearable EEG system for evaluating driver vigilance. *IEEE Trans. Biomed. Circuits Syst.* 8, 165–176. doi: 10.1109/TBCAS.2014.2316224
- Lopez-Gordo, M. A., Sanchez Morillo, D., and Pelayo Valle, F. (2014). Dry EEG electrodes. *Sensors* 14, 12847–12870. doi: 10.3390/s140712847
- Makeig, S., Bell, A., Jung, T.-P., and Sejnowski, T. J. (1995). Independent component analysis of electroencephalographic data. *Adv. Neural Inform. Process. Syst.* 8, 145–151.
- Mehreen, A., Anwar, S. M., Haseeb, M., Majid, M., and Ullah, M. O. (2019). A hybrid scheme for drowsiness detection using wearable sensors. *IEEE Sens. J.* 19, 5119–5126. doi: 10.1109/JSEN.2019.2904222
- Molina, E., Correa, Á., Sanabria, D., and Jung, T. P. (2013). “Tonic EEG dynamics during psychomotor vigilance task,” in *Proceedings of the 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, 1382–1385. doi: 10.1109/NER.2013.6696200
- Molina, E., Sanabria, D., Jung, T.-P., and Correa, A. (2019). Electroencephalographic and peripheral temperature dynamics during a prolonged psychomotor vigilance task. *Accid. Anal. Prevent.* 126, 198–208. doi: 10.1016/j.aap.2017.10.014
- Müller-Putz, G., and Wriessnegger, S. C. (2021). “Electroencephalography and brain–computer interfaces,” in *Neuroprosthetics and Brain-Computer Interfaces in Spinal Cord Injury* (New York, NY: Springer), 71–103. doi: 10.1007/978-3-030-68545-4_3
- Neigel, A. R., Claypoole, V. L., Smith, S. L., Waldfogle, G. E., Fraulini, N. W., Hancock, G. M., et al. (2020). Engaging the human operator: a review of the theoretical support for the vigilance decrement and a discussion of practical applications. *Theor. Issues Ergon. Sci.* 21, 239–258. doi: 10.1080/1463922X.2019.1682712
- Noreika, V., Georgieva, S., Wass, S., and Leong, V. (2020). 14 challenges and their solutions for conducting social neuroscience and longitudinal EEG research with infants. *Infant Behav. Develop.* 58, 101393. doi: 10.1016/j.infbeh.2019.101393

- Parasuraman, R., Warm, J. S., and See, J. E. (1998). "Brain systems of vigilance," in *The Attentive Brain* (Cambridge, MA: The MIT Press), 221–256.
- Roman-Gonzalez, A. (2012). "Eeg signal processing for BCI applications," in *Human-Computer Systems Interaction: Backgrounds and Applications 2* (New York, NY: Springer), 571–591. doi: 10.1007/978-3-642-23187-2_36
- Roy, R. N., Charbonnier, S., Campagne, A., and Bonnet, S. (2016). Efficient mental workload estimation using task-independent EEG features. *J. Neural Eng.* 13, 26019. doi: 10.1088/1741-2560/13/2/026019
- Ruffini, G., Dunne, S., Fuentemilla, L., Grau, C., Farres, E., Marco-Pallarés, J., et al. (2008). First human trials of a dry electrophysiology sensor using a carbon nanotube array interface. *Sens. Actuat. A Phys.* 144, 275–279. doi: 10.1016/j.sna.2008.03.007
- Saeed, S. M. U., Anwar, S. M., Majid, M., and Bhatti, A. M. (2015). "Psychological stress measurement using low cost single channel EEG headset," in *Proceedings of the 2015 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*. IEEE, 581–585. doi: 10.1109/ISSPIT.2015.7394404
- Schaffer, C. (1993). Selecting a classification method by cross-validation. *Mach. Learn.* 13, 135–143.
- Schultze-Kraft, M., Dähne, S., Gugler, M., Curio, G., and Blankertz, B. (2016). Unsupervised classification of operator workload from brain signals. *J. Neural Eng.* 13, 36008. doi: 10.1088/1741-2560/13/3/036008
- Schwarz, A., Escolano, C., Montesano, L., and Müller-Putz, G. R. (2020). Analyzing and decoding natural reach-and-grasp actions using gel, water and dry EEG systems. *Front. Neurosci.* 14, 1–17. doi: 10.3389/fnins.2020.00849
- Sciaraffa, N., Borghini, G., Di Flumeri, G., Cincotti, F., Babiloni, F., and Aricò, P. (2021). Joint analysis of eye blinks and brain activity to investigate attentional demand during a visual search task. *Brain Sci.* 11, 562. doi: 10.3390/brainsci11050562
- Sciaraffa, N., Flumeri, G., Di, Germano, D., Giorgi, A., Florio, A., Di, Borghini, G., et al. (2022). Validation of a light EEG-based measure for real-time stress monitoring during realistic driving. *Brain Sci.* 12:304. doi: 10.3390/brainsci12030304
- Sebastiani, M., Di Flumeri, G., Aricò, P., Sciaraffa, N., Babiloni, F., and Borghini, G. (2020). Neurophysiological vigilance characterisation and assessment: laboratory and realistic validations involving professional air traffic controllers. *Brain Sci.* 12, 304. doi: 10.3390/brainsci10010048
- Seo, S.-H., and Lee, J.-T. (2010). *Stress and EEG. Convergence and Hybrid Information Technologies*. Vienna: IntechOpen. doi: 10.5772/9651
- Skoluda, N., Strahler, J., Schlotz, W., Niederberger, L., Marques, S., Fischer, S., et al. (2015). Intra-individual psychological and physiological responses to acute laboratory stressors of different intensity. *Psychoneuroendocrinology* 51, 227–236. doi: 10.1016/j.psyneuen.2014.10.002
- Somers, B., Francart, T., and Bertrand, A. (2018). A generic EEG artifact removal algorithm based on the multi-channel Wiener filter. *J. Neural Eng.* 15, 36007. doi: 10.1088/1741-2552/aaac92
- Stanton, N. A., and Young, M. S. (2000). A proposed psychological model of driving automation. *Theor. Issues Ergon. Sci.* 1, 315–331. doi: 10.1080/14639220052399131
- Tautan, A.-M., Mihajlovic, V., Chen, Y.-H., Grundlehner, B., Penders, J., and Serdijn, W. A. (2014). "Signal quality in dry electrode EEG and the relation to skin-electrode contact impedance magnitude," in *Biodevices*, 12–22.
- Tăutan, A. M., Mihajlović, V., Chen, Y. H., Grundlehner, B., Penders, J., and Serdijn, W. (2014). "Signal quality in dry electrode EEG and the relation to skin-electrode contact impedance magnitude," in *BIODEVICES 2014 7th International Conference on Biomedical Electronics and Devices, Proceedings; Part of 7th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2014*, 12–22.
- Topor, M., Opitz, B., and Dean, P. J. A. (2021). In search for the most optimal EEG method: a practical evaluation of a water-based electrode EEG system. *Brain Neurosci. Adv.* 5, 239821282110536. doi: 10.1177/23982128211053698
- Toyama, S., Takano, K., and Kansaku, K. (2012). A non-adhesive solid-gel electrode for a non-invasive brain-machine interface. *Front. Neurol.* 3, 114. doi: 10.3389/fneur.2012.00114
- Volosyak, I., Valbuena, D., Malechka, T., Peuscher, J., and Gräser, A. (2010). Brain-computer interface using water-based electrodes. *J. Neural Eng.* 7, 066007. doi: 10.1088/1741-2560/7/6/066007
- Vourvopoulos, A., Niforatos, E., Hlinka, M., Skola, F., and Liarokapis, F. (2017). "Investigating the effect of user profile during training for BCI-based games," in *Proceedings of the 2017 9th International Conference on Virtual Worlds and Games for Serious Applications (VS-GAMES)*. IEEE, 117–124. doi: 10.1109/VS-GAMES.2017.8056579
- Webster, J. G. (2009). *Medical Instrumentation: Application and Design*. Hoboken, NJ: Wiley.
- Wei, C. S., Wang, Y., Te, Lin, C. T., and Jung, T. P. (2018). Toward drowsiness detection using non-hair-bearing EEG-based brain-computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 400–406. doi: 10.1109/TNSRE.2018.2790359
- Wetherell, M. A., and Sidgreaves, M. C. (2005). Secretory immunoglobulin-A reactivity following increases in workload intensity using the Defined Intensity Stressor Simulation (DISS). *Stress Health J. Int. Soc. Investig. Stress* 21, 99–106. doi: 10.1002/smi.1038
- Wickens, C. D. (2008). Multiple resources and mental workload. *Hum. Fact.* 50, 449–455. doi: 10.1518/001872008X288394
- Wilkinson, R. T., and Houghton, D. (1982). Field test of arousal: a portable reaction timer with data storage. *Hum. Fact.* 24, 487–493.
- Yerkes, R. M., and Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Punish. Issues Exp.* 1908, 27–41.
- 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
- Zander, T. O., and Kothe, C. (2011). Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. *J. Neural Eng.* 8, 25005. doi: 10.1088/1741-2560/8/2/025005
- Zander, T. O., Lehne, M., Ihme, K., Jatzev, S., Correia, J., Kothe, C., et al. (2011). A dry EEG-system for scientific research and brain-computer interfaces. *Front. Neurosci.* 5, 1–10. doi: 10.3389/fnins.2011.00053
- Zander, T. O., Phypa, T., and Jatzev, S. (2009). Detecting affective covert user states with passive. *Brain-Computer Interf.* 2, 54. doi: 10.1109/ACII.2009.5349456
- Zander, T. O., Shetty, K., Lorenz, R., Leff, D. R., Krol, L. R., Darzi, A., et al. (2017). Automated task load detection with electroencephalography: towards passive brain-computer interfacing in robotic surgery. *J. Med. Robot. Res.* 2, 1750003. doi: 10.1142/S2424905X17500039
- Zhou, Y., Huang, S., Xu, Z., Wang, P., Wu, X., and Zhang, D. (2021). Cognitive workload recognition using EEG signals and machine learning: a review. *IEEE Trans. Cogn. Develop. Syst.* 8920, 1. doi: 10.1109/TCDS.2021.3090217

Conflict of Interest: NS, GD, DG, AG, AD, GB, AV, VR, FB, and PA were employed by BrainSigns Srl.

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.

Copyright © 2022 Sciaraffa, Di Flumeri, Germano, Giorgi, Di Florio, Borghini, Vozzi, Ronca, Babiloni and Aricò. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



OPEN ACCESS

EDITED BY

Shengzhi Du,
Tshwane University of Technology,
South Africa

REVIEWED BY

Ahmad Mayeli,
University of Pittsburgh, United States
Joanne Yip,
Hong Kong Polytechnic University,
Hong Kong SAR, China
Polina Zioga,
University of Stirling, United Kingdom
Juan Antonio Barrios Heredero,
Miguel Hernández University of Elche,
Spain

*CORRESPONDENCE

Claudia Krogmeier
ckrogmei@purdue.edu

SPECIALTY SECTION

This article was submitted to
Brain-Computer Interfaces,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 25 February 2022

ACCEPTED 21 July 2022

PUBLISHED 12 August 2022

CITATION

Krogmeier C, Coventry BS and
Mousas C (2022) Frontal alpha
asymmetry interaction with an
experimental story EEG
brain-computer interface.
Front. Hum. Neurosci. 16:883467.
doi: 10.3389/fnhum.2022.883467

COPYRIGHT

© 2022 Krogmeier, Coventry and
Mousas. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License](#)
(CC BY). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Frontal alpha asymmetry interaction with an experimental story EEG brain-computer interface

Claudia Krogmeier^{1*}, Brandon S. Coventry^{2,3} and
Christos Mousas¹

¹Department of Computer Graphics Technology, Purdue University, West Lafayette, IN, United States, ²Department of Biomedical Engineering, University of Wisconsin–Madison, Madison, WI, United States, ³Wisconsin Institute for Translational Neuroengineering, University of Wisconsin–Madison, Madison, WI, United States

Although interest in brain-computer interfaces (BCIs) from researchers and consumers continues to increase, many BCIs lack the complexity and imaginative properties thought to guide users toward successful brain activity modulation. We investigate the possibility of using a complex BCI by developing an experimental story environment with which users interact through cognitive thought strategies. In our system, the user's frontal alpha asymmetry (FAA) measured with electroencephalography (EEG) is linearly mapped to the color saturation of the main character in the story. We implemented a user-friendly experimental design using a comfortable EEG device and short neurofeedback (NF) training protocol. In our system, seven out of 19 participants successfully increased FAA during the course of the study, for a total of ten successful blocks out of 152. We detail our results concerning left and right prefrontal cortical activity contributions to FAA in both successful and unsuccessful story blocks. Additionally, we examine inter-subject correlations of EEG data, and self-reported questionnaire data to understand the user experience of BCI interaction. Results suggest the potential of imaginative story BCI environments for engaging users and allowing for FAA modulation. Our data suggests new research directions for BCIs investigating emotion and motivation through FAA.

KEYWORDS

brain-computer interfaces, electroencephalography, neurofeedback, affective computing, story, avant-garde, frontal alpha asymmetry, inter-subject correlations

1. Introduction

Despite the promise of brain-computer interfaces (BCIs) to create personalized and exciting experiences for users (Robinson et al., 2020), BCIs are considered “not ready” for use (Cattan, 2021), as aesthetics in BCI experiences are often not a priority, among other reasons. Thus far, BCIs are not commercially successful (Kerous et al., 2018), and present numerous challenges for researchers (Saha et al., 2021) such as simple interfaces which may not be engaging for users (Cohen et al., 2016). While there exist multiple

types of BCIs, neurofeedback-based BCIs are the most popular (Kerous et al., 2018). Neurofeedback (NF) allows participants to learn to control unconscious brain activity by perceiving feedback concerning their brain activity. Feedback is often visual and simple, such as a thermometer bar (Johnston et al., 2011; Robinson et al., 2020).

While clinical NF often involves extensive training protocols for the purposes of understanding long-term psychological outcomes, BCI research harnesses NF techniques within relatively short training sessions, with the primary goal of examining novel, dynamic interaction methods for participants (Sitaram et al., 2017; Charles et al., 2020). In this study, we develop and investigate an affective BCI, which takes as input correlates of user emotion and motivation through frontal alpha asymmetry (FAA). FAA has been a target measurement in EEG NF because it is relevant to Major Depressive Disorder (MDD) (Zotey et al., 2020), and is often studied in the context of affective and motivational disorders such as MDD and anxiety. Therefore, FAA has been used as input for affective BCIs investigating novel interaction systems for emotion self-regulation. A review of the role of FAA in both emotional and motivational processes can be found at Harmon-Jones and Gable (2018).

FAA can be modulated through NF paradigms in which users learn to regulate brain activity through operant conditioning, despite a lack of volitional control (Rosenfeld et al., 1995; Aranyi et al., 2015b, 2016). Additionally, FAA can potentially serve as a biomarker to study and modulate approach motivation and affect (Briesemeister et al., 2013). In this paper, we detail our developed experimental story BCI which is designed with the goal of stimulating the imagination of the user in order to prompt novel thought content strategies for brain activity modulation. We report our experimental design and analysis, which is inspired by affective BCI work using FAA by Aranyi et al. (2016). We provide a discussion of our findings in order to inform affective BCI protocol design using complex stimuli with the broader goal of investigating user-friendly, higher complexity BCIs.

Our BCI examines participant NF interaction success within an experimental story BCI FAA NF protocol. The experimental story was designed to allow participants to exercise their imagination in order to explore mental strategies for increasing FAA. We aim to understand the efficacy of the developed BCI in allowing participants to increase their FAA score through the examination of the following research questions:

- **RQ1:** Can participants successfully engage with the developed experimental story BCI by increasing FAA?
 - We hypothesize that some, but not all participants will successfully increase FAA, and that successful blocks will be characterized by large effect sizes.

- **RQ2:** Will successful blocks be characterized by an increase in left frontal cortical activity?

- We hypothesize that increased FAA will be the result of increased left frontal cortical activity rather than decreased right frontal cortical activity.

- **RQ3:** Will participant cognitive thought strategies be associated with NF interaction success?

- We hypothesize that successful participants will primarily use direct strategies while unsuccessful participants will primarily use indirect strategies to modulate brain activity.

FAA scores, measured in real-time with electroencephalography (EEG), serve as the input signal in our BCI system. For the participant, visual feedback concerning FAA is mapped to the color saturation of the main character in the story environment. Participants engaged with eight distinct story segments using imagination-based instructions for thought content strategies. Greater FAA scores increased the main character's color saturation, while lower FAA scores decreased color saturation of the main character in the dream-like story environment.

1.1. BCI: Frontal alpha asymmetry (FAA) and EEG

Frontal alpha asymmetry (FAA) refers to a difference in brain activity between the right and the left prefrontal cortices of the brain, and can be measured with a difference score between corresponding right and left electrode sites on an EEG device. Multiple studies have indicated that greater left relative to right prefrontal cortical activity is associated with increased approach motivation, defined as an organism's tendency to approach or expend energy in order to go toward stimuli rather than away (Harmon-Jones and Gable, 2018). Additionally, greater left relative to right prefrontal cortical activity has been associated with better emotion self-regulation ability, increased positive emotions, as well as reduced depressive symptoms (Cohen et al., 2016; Quaedflieg et al., 2016; Harmon-Jones and Gable, 2018). In affective computing, determining new ways in which humans can more directly control emotions is central to the advancement of humanity generally (Cavazza et al., 2014b), as emotional processing is diminished within numerous mental health disorders. Thus, FAA has been studied within numerous NF and BCI paradigms (Peeters et al., 2014; Aranyi et al., 2016; Mennella et al., 2017) as well as applications aiming to alleviate depression symptomology (Kelley et al., 2017).

Our objective is to investigate the efficacy of our developed BCI to facilitate FAA modulation for the user. As BCI research often emphasizes technical aspects over both human interaction and user guidance, we develop our BCI system with a focus on the cybernetic paradigm “human in the loop,” as described by Kosmyna and Lécuyer (2017). For our study, we use the Emotiv Epoc X¹ EEG headset. This headset is minimally intrusive, quick to set up, and affordable enough to be selected by consumers; perhaps why most BCI studies incorporate a consumer grade device (Kerous et al., 2018).

1.2. Experimental storytelling techniques

While there is no concrete definition of “avant-garde,” there exist certain attributes of both film and game works that are considered characteristics which can often, but not always, be found within the genre (Taberham, 2018). While techniques within avant-garde films have been identified and described after their creation, we use many experimental film making techniques seen in avant-garde works within our developed experimental story BCI, seeking to use these identified techniques for the BCI experience, rather than develop a work defined as “avant-garde.” Techniques we employed include rejection of linear narrative, spatio-temporal discontinuity, lack of causal logic, and prominent stylization. The primary goal of avant-garde works is often to provide novel psychological experiences for viewers, in which viewers must exert energy and creative leaps of the imagination to think about and experience the work (Koenitz, 2017; Taberham, 2018). This exercise of the imagination is central to avant-garde works (Taberham, 2018). Therefore, our goal was to explore experimental film-making techniques identified within previously developed avant-garde works, with the goal of stimulating the participant’s imagination during NF interaction. Similar to Zioga et al. (2018), we believe that interdisciplinary research such as this study will allow for creative, new approaches to understanding NF interaction in BCI experiences which may help understand these experiences in real-world settings.

By experiencing the experimental BCI story environment, participants may explore cognitive thought content strategies for brain activity modulation more freely, as creativity and imagination may be necessary to process the story environment. We investigate our assumption that an experimental story BCI may influence user mental strategies for BCI engagement positively by evaluating participant BCI engagement success through FAA measurements. We investigate commonalities in user neural engagement through inter-subject correlations (ISC) of EEG data as exploratory analyses as well, and obtain self-reported questionnaire data in order to provide a more complete picture of the user experience during BCI engagement.

¹ <https://www.emotiv.com/epoc-x/>

2. Related work

Since the 1960s (Nijholt, 2015), artists have advanced the field of BCIs through the creation of applications within real-life contexts (Zioga et al., 2018). Previous research has explored multi-brain activity amongst users through art installation work (Mori, 2003; Albu, 2020), BCI collaborative control of music (Le Groux et al., 2010), and mixed media performances (Zioga et al., 2018); See Wadeson et al. (2015) for more earlier artistic BCI work. Despite many previous, creative BCIs which provide participants with imaginative, playful, as well as collaborative experiences using brain activity, BCI research investigating NF interaction success within more engaging BCI experiences is limited.

Many NF protocols investigated through user studies involve simple environments such as those which utilize status bars, single auditory tones, shape changes (Cavazza et al., 2017) or color changes (Cohen et al., 2016) for feedback concerning the user’s brain activity. Moving cubes (Berger and Davelaar, 2018), histograms (Mennella et al., 2017), and box-plot meters (Quaedflieg et al., 2016) have also been used to visualize brain activity for participants. Recent work investigating the combined usage of EEG and functional magnetic resonance imaging (fMRI) NF for emotion self-regulation additionally used colored height bars to visualize two signals from EEG and two signals from fMRI (Zotev et al., 2020), as well as colored height bars in combination with pictures of positive autobiographical memories (Dehghani et al., 2020). In a study by Lackner et al. (2016), brain activity was visualized through the movement of a ball, which changed from blue to yellow with the participant’s alpha band activity. Participants also saw a happy or sad smiley face after each run. Research suggests that improving the complexity of NF environments could lead to increased NF success (Cohen et al., 2016).

Cohen et al. (2016) compared EEG NF success in emotion down-regulation using a virtual animated scenario to using a simple thermometer 2D feedback system. The authors determined that the more complex, animated scenario was not only more effective for NF learning, but was also more motivating, engaging, and allowed for a greater NF learning transferability to unfamiliar environments. Berger and Davelaar (2018) examined attentional control in their BCI system, and determined that participants had a higher learning rate for increasing attention in their 3D environment than in their 2D environment. Gruzelier et al. (2010) determined greater NF learning for participants in their Cave Automatic Virtual Environment (CAVE) than for participants experiencing a screen-based rendition of the same environment. Although these studies examined varying attributes of environments with higher complexity, greater interface complexity appears to have contributed to increased NF success across different platforms.

While literature concerning affective BCIs using FAA is limited (Aranyi et al., 2015b), there exist several BCI FAA studies which utilize stories to enhance participant motivation and success through an encouragement of the imagination. Cavazza et al. (2014b) examined the use of an interactive narrative for affective interaction. In their BCI, participants were instructed to think positive thoughts in order to support the main character of their story. In their story, a female doctor experiences stressful life events such as patient death, abuse from her boss, and overworking. The participant must mentally support this character by using their thoughts in order for the doctor's outcome to improve throughout the story. Participants' brain activity was mapped to the color of the female doctor, whose color became more saturated as FAA increased. Although a proof of concept study, half of the participants were able to increase FAA with minimal training by mentally supporting the protagonist through empathetic feelings.

Aranyi et al. (2015b) developed a story-based affective BCI investigating FAA using a similar hospital story BCI environment. Participants were instructed to express angry thoughts toward an evil character, as anger has also been associated with increased approach motivation and greater FAA scores despite being a negative emotion. Brain activity was mapped to the alpha channel of the evil character. In order to prime participants to feel anger toward the evil character, this character was depicted abusing co-workers in the story environment. With greater FAA scores, the evil character became translucent, as if he was disappearing from the scene. With lower FAA scores, the evil character became more opaque, remaining in the story. Despite minimal training, participants were able to successfully engage with the BCI by increasing FAA through anger expression.

In a later study also investigating FAA, Aranyi et al. (2016) developed an affective BCI which mapped FAA scores to facial expressions of a virtual agent (character). Participants were instructed to express positive thoughts toward the agent. With increased FAA, the agent would show a positive, happy expression, and with decreased FAA, the agent would exhibit a neutral expression or look away from the participant. The authors determined that a majority of participants could successfully increase FAA through their affective BCI paradigm. For their BCI, the authors developed a short, user-friendly NF protocol designed to control for differences in FAA both *between* different NF blocks, and *across* different participants (Aranyi et al., 2016). Our methodology is largely based upon the NF protocol developed by Aranyi et al. (2016), as we seek to investigate our BCI within a user-friendly NF protocol.

We investigate BCI complexity through the development and evaluation of an experimental story environment BCI, utilizing simple visual feedback in the form of color saturation, as this visual feedback has been used in previous BCI studies. Considering that multi-component BCIs have allowed for greater NF success than single component BCIs (Jensen et al.,

2013), our BCI consists of eight distinct story segments with which participants could interact through brain activity. With this study, our primary goal was to investigate an imaginative, complex BCI environment by providing participants with an experimental story experience and imagination-based instructions for engagement.

3. Materials and methods

3.1. Experimental story BCI system overview

Our BCI system was developed in the Unity game engine using the Emotiv plugin. To provide participants with real-time visual feedback of their brain activity (FAA), FAA scores were calculated by subtracting the natural log-transformed alpha power of the left electrode (F3) from the natural log-transformed alpha power of the right electrode (F4) ($\ln[Right] - \ln[Left]$) at a 2 Hz sampling rate, to match prior research by Aranyi et al. (2016). For *post-hoc* analyses, FAA scores were calculated after preprocessing the raw EEG data in EEGLAB (Delorme and Makeig, 2004). Our *post-hoc* analyses determined participant success rate in increasing FAA from *View* to *Engage* components.

Our protocol is based on that of Aranyi et al. (2016), who developed a similarly short NF protocol investigating FAA. In our protocol, participants completed eight different story blocks. Each story block consisted of a 30 s *View* component, and a 30 s *Engage* component. Each of the eight *View* and *Engage* segment pairs were separated by a 10 s *Rest* component for a total of eight *Rest* components. During *Rest*, participants were instructed to relax and try to minimize mental wandering. The timing of our *Rest* and *Prompt* components matches that of Aranyi et al. (2016). While Aranyi et al. (2016) presented *Engage* and *View* components for 40 s, our protocol differs in that we presented these components for 30 s each. This decision was made in order to both match the eight NF block protocol of Aranyi et al. (2016), while keeping the BCI experience under 15 min for participants as mentioned in Section 3.3.

During *View*, participants were instructed to count backwards. Participants counted backwards by 2 s silently starting at 500 (500, 498, 496, etc.) during each of the eight *View* components. The number by which to count backwards was determined through our pre-testing. It was determined that counting backwards by two was adequate to maintain attention while not being so difficult as to distract and frustrate participants. Counting backwards was used to control for unwanted thoughts, emotions, and mental processes (Aranyi et al., 2016).

During *Engage* components, participants were instructed to use their thoughts to interact with the main character in the story, as described in more detail in the feedback mapping subsection of this paper (see Section 3.3). A 3 s prompt screen

was presented to participants before both *View* and *Engage* components to indicate which component would be presented next. Participants were made aware that they would receive no visual feedback concerning their brain activity during *View* components, and would only see visual feedback during *Engage* components. Participants described mental strategies they used during the experiment on the questionnaire. One of the eight story segments of our NF protocol is shown in [Figure 1](#).

3.1.1. Story blocks

We developed the story using the following design techniques: spatio-temporal discontinuity, rejection of linear narrative, abstraction of main character and narrative, and prominent stylization. Live action video footage was filmed and used for background textures within the story environment in order to create a dream-like mood. A young female 3D model was downloaded from Adobe Mixamo² and used as the main character, while the background character, an adult male 3D model, was downloaded from the Microsoft Rocketbox Avatar library, MoveBox ([Gonzalez-Franco et al., 2020](#)). A shader program downloaded from the Unity asset store was applied to both characters in order to distort spatial logic concerning their forms. During *View* components, the main character and the background character were unlit (no colors), with the shader effect applied consistently throughout the *View* component. At the beginning of each *Engage* component, the main character faded out of the shader effect, returning to the original shape of the 3D model, to clearly indicate the start of an *Engage* component. During *Engage* components, participants were able to use their thoughts to increase the main character's color saturation. Characters were animated with very slow animations during both *View* and *Engage* components. Because increasing color saturation has been used previously in an FAA NF BCI ([Cavazza et al., 2014b](#)), as well as the manipulation of character transparency ([Aranyi et al., 2015b](#)), we selected an interaction mapping modality known to produce NF interaction success. In this way, we could examine the influence of an experimental story context on NF interaction success while employing a known visual feedback modality. [Jensen et al. \(2013\)](#) suggests examining how a combination of components might affect the user's imagination for usage in neurofeedback. All eight story blocks, consisting of both *View* and *Engage* components are shown in [Figure 2](#).

Concerning the spatiality of the story environment, 3D models in certain story blocks were positioned in ways that could not be possible in reality. For example, a set of chairs and a table were positioned diagonally, with a 45° tilt downwards, as the characters hovered above the environment in Story Block #3. In Story Block #5, the background character hovers well above the ground. These story blocks can be seen in [Figure 3](#).

² <https://www.mixamo.com/>

Lacking a clear chronology of events, the story blocks were presented in the same order for all participants. Previously developed affective story BCIs have incorporated environments with negative connotations, with the goal of guiding participants more easily toward mindsets which may approach the target brain activity ([Gilroy et al., 2013](#); [Cavazza et al., 2014a,b](#); [Aranyi et al., 2015b](#)). Our story environment depicted the main character being symbolically watched, controlled, chased, etc., as our goal was to motivate participants to help the main character escape from the bad dream-like environment by using their thoughts. In pre-testing, participants acknowledged that the story was dream-like, open-ended, and interesting, and offered varying ideas concerning its meaning. Therefore, we concluded that the developed story environment was appropriately abstract to allow for a multiplicity of interpretations from participants.

3.2. BCI input

We used the Emotiv Epoc X EEG headset (Emotiv Systems Inc., San Francisco, CA, USA) to collect EEG data from participants. The Emotiv Epoc X is a consumer grade, 14 channel EEG headset which includes electrode sites AF4, AF3, F3, F4, F7, F8, FC5, FC6, O1, O2, P7, P8, T7, and T8. This electrode scheme is based on the international 10-20 system, as shown in [Figure 4](#). Two additional electrodes served as reference and ground; the electrode located at the M1 site (Driven Right Leg [DRL]) functioned as an absolute voltage reference while the electrode located at the M2 site (Common Mode Sense [CMS]) was used for feedback noise cancellation. The electrodes are Ag/AgCl sensors which contain felt pads. Felt pads were fully soaked in saline solution prior to inserting into each electrode compartment in order to ensure the best conductance between the participant's scalp and the sensor. EEG data was recorded with a 256 Hz sampling rate and was filtered online using a built-in digital 5th order Sinc filter with a bandwidth of 16–43 Hz, with notch filters at 50 and 60 Hz.

Consumer-grade EEG data may provide researchers with more ecologically valid results, as the headset is both minimally obtrusive for participants, and has demonstrated EEG data consistent with conventional EEG recordings ([Le et al., 2020](#)). According to [Gapen et al. \(2016\)](#), EEG NF has made very little impact in clinical care despite being used as a clinical intervention for more than 30 years. Considering that wireless, portable and small EEG systems which can be used outside traditional laboratory environments may increase clinical relevance ([Enriquez-Geppert et al., 2017](#)), we use such an EEG device with the goal of investigating NF in a more real-world setting.

F3 and F4, along with F7 and F8, are the most commonly investigated electrode pairs in FAA research ([Smith et al., 2017](#); [Kuper et al., 2019](#); [David et al., 2021](#)). To maintain consistency between this study and our previous research, we

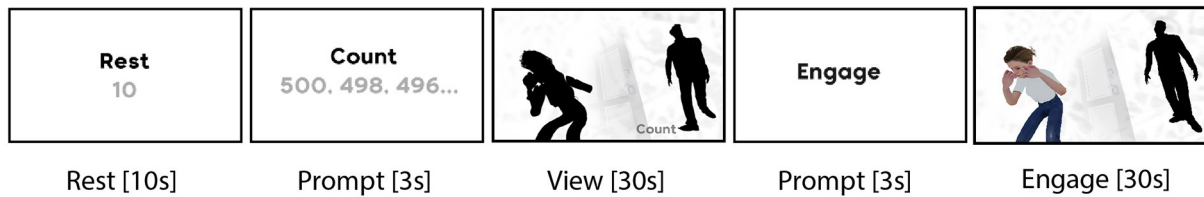


FIGURE 1

One block of our NF protocol. Components include (A) Rest [10 s], (B,D) Prompt [3 s], (C) View [30 s], and (E) Engage [30 s].

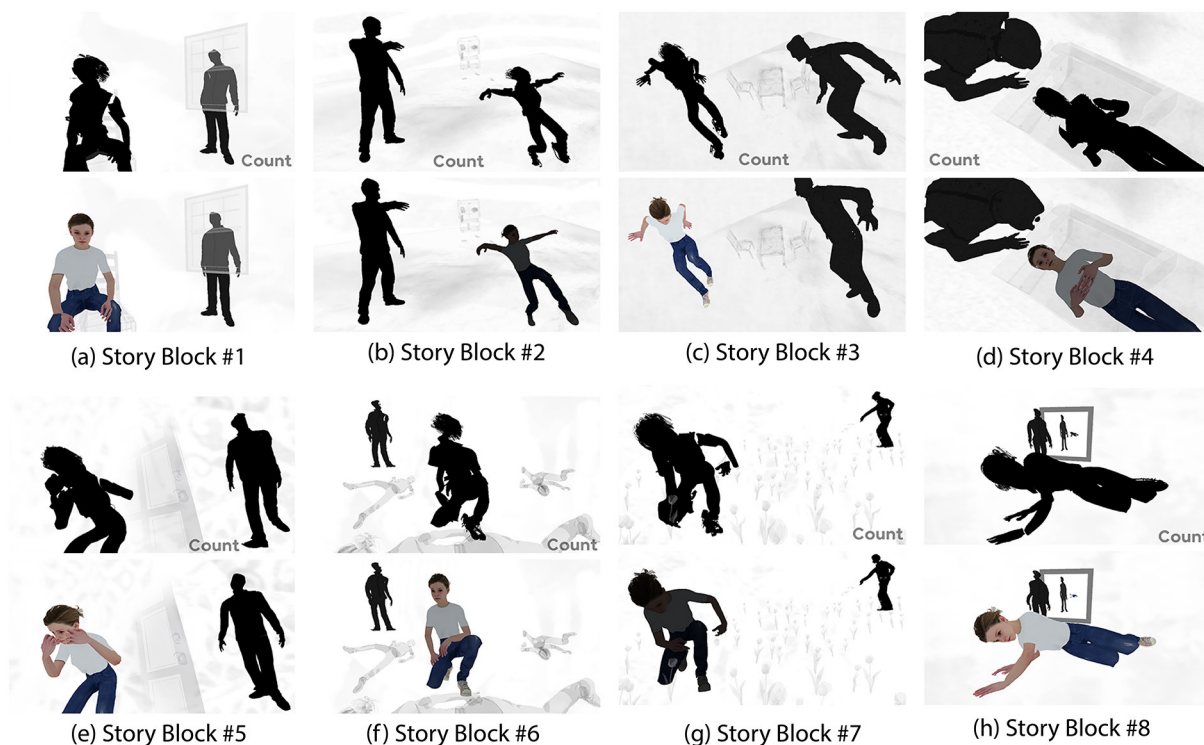


FIGURE 2

All eight story blocks, with View and Engage components. 2b and 2g demonstrate FAA which has not crossed the threshold for visual feedback, therefore, the 3d model remains dark. (a) Story block #1, (b) Story block #2, (c) Story block #3, (d) Story block #4, (e) Story block #5, (f) Story block #6, (g) Story block #7, (h) Story block #8.

selected electrode sites F3 and F4 for FAA calculation to provide participants with real-time visual feedback concerning their brain activity.

3.3. Participants, procedure, and feedback mapping

Data was collected from 23 participants, all students at Purdue University. Due to excessive eye closure, movements, or EEG signal capture malfunctions, four participants were removed from the analysis, resulting in 19 participants (five female and 14 male; age: $M = 19.05$, $SD = 1.02$; age

range: 18–22). With the exception of one participant who was left-handed, all participants were right-handed. Only one participant was currently undergoing psychiatric treatment. These participants were not identified as outliers, and were therefore included in the analysis. All participants had normal or corrected-to-normal vision. The study was approved by Purdue University's institutional review board (IRB), and participants provided written consent before participation. Similar to the BCI cinematic experience created by Pike et al. (2016), participants viewed the BCI on a large screen, in a quiet room. Participants were seated in a comfortable chair in order to minimize motion artifacts, and viewed the monitor from a comfortable viewing distance. The study was completed during the week over a

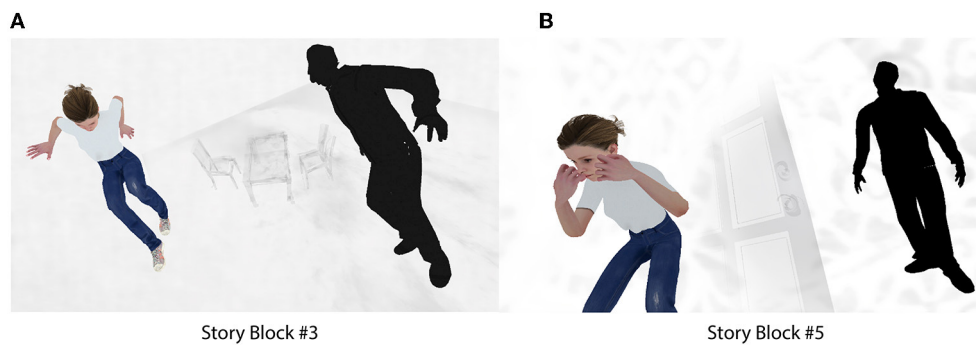


FIGURE 3
Spatio-temporal discontinuity in our story environment. (A) Story block #3, (B) Story block #5.

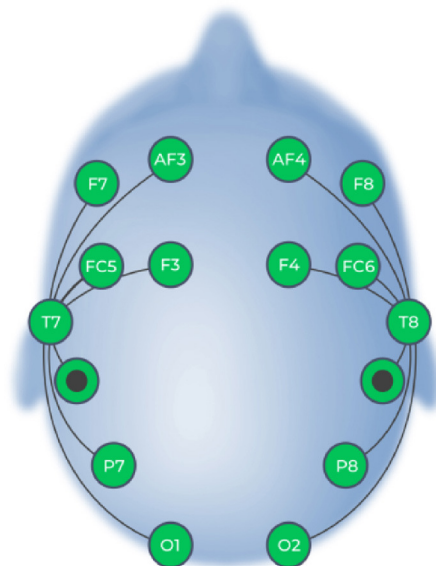


FIGURE 4
Electrodes in the Emotiv Epoc X System (image courtesy of Emotiv).

three week period in late fall. Participants took 35–45 min to complete the entire study, depending on how long it took them to finish the questionnaire after the experiment. This time included EEG preparation, participant instructions, a practice round, the experiment, and questionnaire completion. The total time spent engaging with the BCI during the experiment was 13.5 min, as it has been suggested that BCI tasks longer than 15–20 min can greatly contribute to participant fatigue and, therefore, diminishing EEG data quality (Aranyi et al., 2015a). Figure 5 shows a participant in the experiment, and the devices used in the study.

The researcher first wiped the participant's lateral forehead and mastoid regions with alcohol to ensure similarly clean

recordings for each participant. After the EEG headset was positioned on the head, the researcher adjusted participant hair until the Emotiv system indicated 100% EEG signal quality (all electrode indicators were green, which indicated sufficient conductance levels). To increase participant compliance with refraining from movements during the study, the researcher asked participants to try clenching their jaw, frowning, talking, and moving their limbs and face, and showed participants the noise introduced into their EEG signal during movements such as these. Additionally, the researcher showed participants what the main character would look like with full color saturation. Before the experiment, participants completed one practice block in which they practiced counting backwards during *View* components, and practiced using their thoughts to change color saturation of the main character during *Engage* components. The researcher ensured that participants clearly understood these tasks prior to starting the experiment.

Instructions concerning NF strategies are often vague so as not to constrain participants to strategies which may not be effective for them (Aranyi et al., 2016). During pre-testing, participants expressed confusion with our initial instructions. Therefore, a new set of instructions was developed. These instructions were more detailed, while maintaining an open-endedness. Defined as the ability to mentally reconstruct new information, sensations, and objects, imagination (Szczelkun, 2018) was central to our participant instructions, provided below.

In the experiment, you will observe the girl in her dream. Your goal is to help her escape from her dream. During Engage components, you may think about new objects, sensations, ideas and possibilities, or new interpretations of the story, which may help the girl escape from this dream environment. You may imagine interacting with the girl. You may try using positive thoughts. You may explore other strategies for changing the main character's color as well.

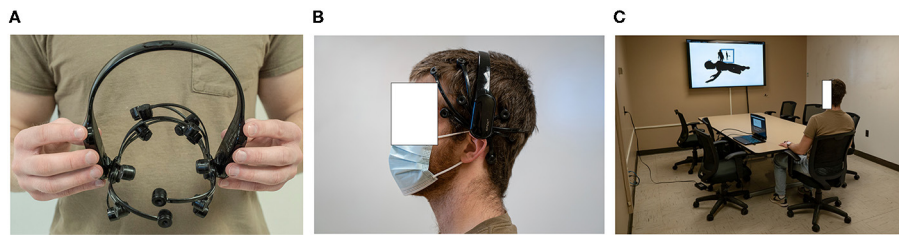


FIGURE 5

The user and devices in our BCI system. (A) Emotiv Epoc X EEG headset. (B) User wearing the headset. (C) The BCI room set-up.

F3 and F4 Electrodes



FIGURE 6

BCI system overview.

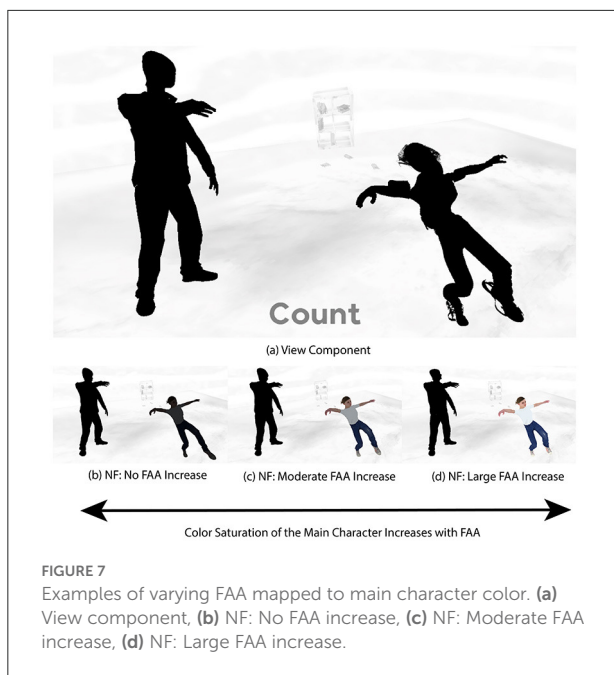
Following work by Aranyi et al. (2016), we defined the brain activity (FAA) threshold for visual feedback based on guidance for EEG-based FAA in NF paradigms (Rosenfeld et al., 1995). FAA values from each *View* component were used to determine the minimum and maximum FAA values necessary for visual feedback for each corresponding *Engage* component. Before calculating the minimum and maximum visual feedback threshold points, outliers that were three standard deviations higher or lower than the mean of the list of FAA values stored during *View* were removed from the list of values, in order to remove extreme values indicative of participant movements. Next, the minimum value for visual feedback during the corresponding *Engage* component was defined as the mean of the FAA values (excluding outliers) during the preceding *View* component plus 1.28 times their standard deviation. This threshold requirement for minimum values determined that visual feedback during NF would result in no feedback for 90% of the values from the *View* component, and instead show visual feedback for only the top 10% of values from the *View* component (Rosenfeld et al., 1995; Aranyi et al., 2016). The maximum value for visual feedback (the highest FAA value included in visual feedback during the *Engage* component) was set as the maximum value recorded during *View*.

FAA values within the minimum and maximum range were linearly mapped to the main character's color saturation values, with the maximum value resulting in the main character being fully colored. Figure 6 shows an overview of our BCI system.

View and *Engage* components were 30 s in length, as determined by our pre-testing. During *Engage* components only, FAA was mapped to the color saturation of the main character, thus allowing the participant to change the color saturation of the main character using their thoughts. Varying color saturation of the main character based on FAA input can be seen in Figure 7. The counting task during each *View* component provided an emotional control for each corresponding *Engage* component (Aranyi et al., 2016). A video showing the developed BCI can be found in our [Supplementary material](#).

3.3.1. EEG preprocessing and analysis

To examine participant success in increasing FAA from *View* to *Engage post-hoc*, we first preprocessed EEG data using EEGLAB (Delorme and Makeig, 2004) within MATLAB (MathWorks Inc.). Data was filtered between 1 Hz (high pass) and 30 Hz (low pass) using EEGLAB's `eegfiltnew` function, following FAA EEG VR research by Kisker et al. (2021). Bad channels were rejected and interpolated. After applying an average reference, Independent Component Analysis (ICA) was run on the data, and eye components with 50% probability were removed. EEGLAB's `spectopo` function was used on the cleaned data to calculate the power spectra in the 8–13 Hz alpha range. We subtracted the natural log-transformed alpha power of the left electrode (F3) from the natural log-transformed alpha power of the right electrode (F4) to calculate FAA: $(\ln[\text{Right}] - \ln[\text{Left}])$ (Smith et al., 2017). FAA was calculated for each *View*



and *Engage* component for each participant. Within each 30 s component, FAA was calculated from 2 s segments to create a distribution of FAA scores.

3.3.2. FAA and alpha power measurements

We examine both FAA measurements as well as alpha power measurements from the left and right electrode sites individually in our study. FAA is calculated by subtracting the natural log of the alpha power from the left electrode site from the natural log of the alpha power of the right electrode site. However, this FAA score does not provide information concerning how much the alpha power from either the left or right electrode site contributed to the FAA score (Smith et al., 2017). Like Aranyi et al. (2016), we wanted to understand the contribution of right and left prefrontal cortical activity to successful FAA scores, as a higher FAA score could be indicative of either increased left cortical activity or decreased right cortical activity (Smith et al., 2017). Therefore, we conducted additional analyses to understand how alpha power on the right and left sides changed from *View* to *Engage* blocks. In this way, we wanted to understand if FAA which increased from *View* to *Engage* did so because of increased left prefrontal cortical activity, or decreased right prefrontal cortical activity. Because alpha power is inversely related to cortical activity (Allen et al., 2004), a higher alpha power from the right electrode would indicate lower cortical activity from the right electrode. Results concerning alpha power from the right and left hemispheres are described in Sections 4.1 and 4.2.

3.3.3. Inter-subject correlations (ISC) analysis

To determine between subject reliability of evoked responses, inter-subject correlations (ISC) were calculated (adapted from Dmochowski et al., 2012; Cohen et al., 2016). Briefly, sample electrode covariances were calculated as:

$$R_{ij} = \frac{1}{N} \sum_t x_i(t)x_j(t), \quad (1)$$

where x_n represents the time series electrical activity recorded on all electrodes of subject n . Within and between subject correlations were then calculated as:

$$R_w = \frac{1}{N} \sum_n R_{nn} \quad (2)$$

$$R_b = \frac{1}{N(N-1)} \sum_i \sum_{j \neq i} \sum_k (x_i(t) - \bar{x}_i)(x_k(t) - \bar{x}_k)^T, \quad (3)$$

where $x_n(t)$ is the measured scalp voltage on channel n and \bar{x}_n is the time average of channel n . The value R_b represents the summation over all cross-covariances of all electrodes of all subjects. Maximal covariances are then calculated as component projections:

$$\text{Component}_i = \frac{v_i^T R_b v_i}{v_i^T R_w v_i}, \quad (4)$$

where v_i is the i^{th} eigenvector of the matrix $R_w^{-1} R_b$. Intersubject correlation is then calculated as:

$$\text{ISC} = \sum_i C_i. \quad (5)$$

In keeping with previous ISC studies (Dmochowski et al., 2012; Cohen et al., 2016), the three largest correlated components are utilized in calculating ISC. Time resolved correlations were formed by calculating ISCs across all electrodes for all subjects for each scene within 1 s windows with an 800 ms overlap between windows.

Spatial distributions of maximal correlation coefficients were calculated using “forward model” analyses (Parra et al., 2005), specifically:

$$A = R_w W (W^T R_w W)^*, \quad (6)$$

where W is the set of linear spatial filters:

$$w_{ij} = \operatorname{argmax}_w \frac{w^T R_{ij} w}{\sqrt{w^T R_{ii} w} \sqrt{w^T R_{jj} w}}, \quad (7)$$

and $*$ designates the Moore-Penrose pseudo-inverse (Ben-Israel and Greville, 2003), a generalization of the matrix inverse which calculates a least-squares best fit *via* singular value

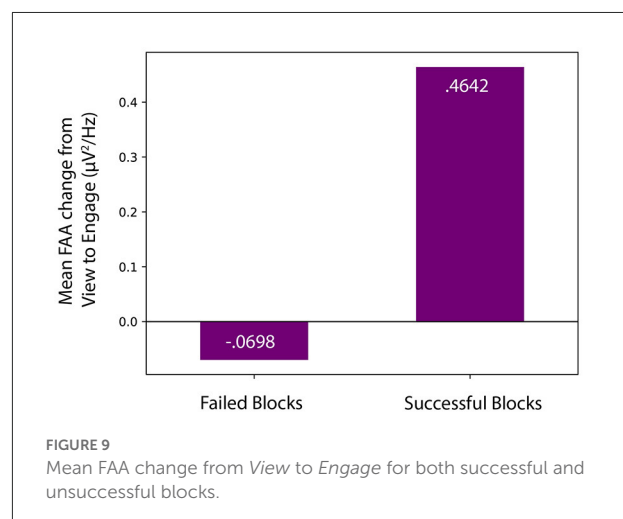
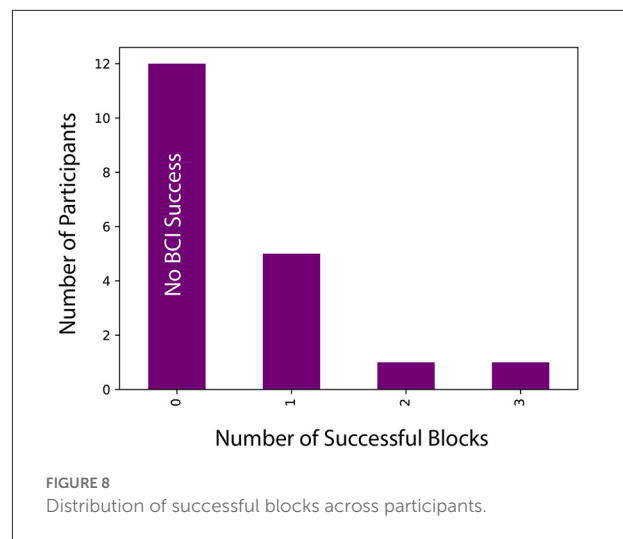
decomposition to calculate a matrix inverse. The Moore-Pensore is used to ensure numerical stability between electrodes with varying levels of activation.

Statistical significance of ISCs was assessed *via* permutation testing (Fischer, 1971). Electrode time series were randomly shuffled in time and empirical p-values reflective of one-tailed type-1 error in rejecting the null hypothesis were calculated from the distribution of time-shuffled correlations after 1,000 iterations. A window of ISC was considered significant if $p < 0.05$ after multiple comparisons correction using the Benjamini-Hochberg procedure to control for the false discovery rate (Benjamini and Hochberg, 1995).

4. Results and discussion

One hundred and fifty-two blocks were identified for analysis (eight blocks for each of the 19 participants). Block success was defined as a statistically significant increase in average FAA during *Engage* components compared to average FAA during *View* components within the same block. For example, we compared the *Engage* and *View* components for Story Block #1, then for Story Block #2, and so on, for each participant individually. Block success was determined using paired samples *t*-tests. Data was inspected using Q-Q plots of the residuals and was normally distributed. Like Le et al. (2020), who also investigated FAA using an Emotiv EEG recording device, we winsorized outliers in the data in order to reduce the influence of extreme scores. We winsorized the data by replacing outlier values with the next minimum or maximum value for that component (as the data included both positive and negative FAA values). Seven out of 19 participants (37%) achieved statistically significant NF success in increasing their FAA score in at least one of the eight blocks, for a total of 10 successful blocks in the experiment. Figure 8 shows the distributions of successful blocks across participants.

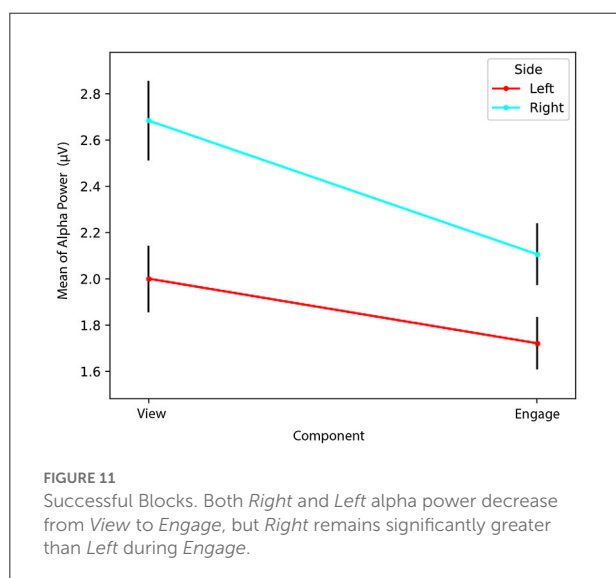
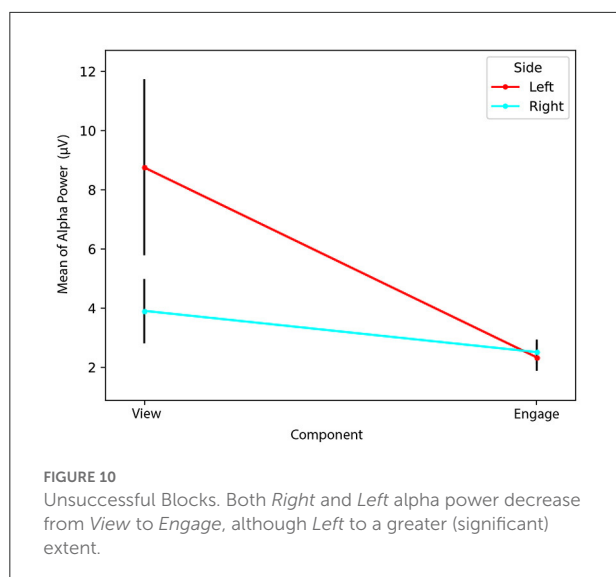
We investigated effect sizes for successful blocks, in which average FAA during *Engage* was statistically significantly greater than average FAA during *View*. Effect sizes for successful participant blocks were calculated using Hedge's *g* using a standard deviation of the difference. The smallest and largest effect sizes detected were 0.477 and 1.27, respectively, with an average effect size of 0.749. Similar to work by Aranyi et al. (2016), our effect size values were negatively skewed, indicative of larger effect sizes for successful blocks. Additionally, we determined that successful blocks were characterized by a large FAA increase ($M = 0.46$, $SD = 0.04$) from *View* to *Engage*, while unsuccessful blocks were characterized by a very small decrease in FAA ($M = -0.07$, $SD = 0.01$), as shown in Figure 9. Despite few successful blocks, our **RQ1** is supported as some participants successfully increased FAA scores from *View* to *Engage*, and successful blocks were characterized by large effect sizes as well as large FAA increases.



4.1. Successful blocks characterized by increased left prefrontal cortical activity from *View* to *Engage*

We used a two-way repeated measures analysis of variance (ANOVA) with average alpha power as the dependent variable, and *Component Type* (*View* or *Engage*) and *Side* (*Left* [F3] or *Right* [F4]) as factors in the analysis. Alpha power which was not normally distributed was transformed following recommendations by Templeton (2011) prior to conducting our parametric analyses. Our simple main effects analysis used Bonferroni confidence interval adjustment.

We found a statistically significant interaction effect between *Side* and *Component Type* [$\Lambda = 0.561$, $F_{(1,9)} = 7.035$, $p = 0.026$, $\eta_p^2 = 0.439$]. Therefore, we conducted simple main effects analyses for both *Side* and *Component Type*. We determined a simple main effect for *Side*, finding that *Left* alpha power ($M = 8.74$, $SD = 9.41$) during *View* was significantly greater



than *Right* alpha power ($M = 3.90, SD = 3.45$) during *View* [$\Lambda = 0.619, F_{(1,9)} = 5.548, p = 0.043, \eta_p^2 = 0.381$]. However, *Left* alpha power ($M = 2.34, SD = 1.49$) did not significantly differ from *Right* alpha power ($M = 2.51, SD = 1.29$) during *Engage* [$\Lambda = 0.960, F_{(1,9)} = 0.373, p = 0.556, \eta_p^2 = 0.040$].

We also found a simple main effect for *Component Type*, determining that *Left* alpha power ($M = 8.74, SD = 9.41$) during *View* was significantly greater than *Left* alpha power ($M = 2.34, SD = 1.49$) during *Engage* [$\Lambda = 0.591, F_{(1,9)} = 6.231, p = 0.034, \eta_p^2 = 0.409$]. However, we found that *Right* alpha power ($M = 3.90, SD = 3.45$) during *View* did not significantly differ from *Right* alpha power ($M = 2.51, SD = 1.29$) during *Engage* [$\Lambda = 0.745, F_{(1,9)} = 3.085, p = 0.113, \eta_p^2 = 0.255$].

Our results indicate greater *Right* cortical activity than *Left* cortical activity during *View*. Because greater relative *Right* than *Left* cortical activity would indicate a lower FAA than if $Left > Right$, it is possible that successful blocks were characterized by a lower starting FAA during *View* compared to *Engage*. Although *Right* cortical activity was greater than *Left* cortical activity during *View*, *Right* and *Left* cortical activity do not differ significantly during *Engage*, as seen in Figure 10. Considering the decrease in alpha power (increase in cortical activity) on the *Left* from *View* to *Engage*, an increase in *Left* cortical activity appears to have contributed to increased FAA, rather than a decrease in *Right* cortical activity from *View* to *Engage*. We consider our **RQ2** partially supported, as an increase in *Left* cortical activity from *View* to *Engage* contributed to increased FAA during *Engage*, despite determining no difference between *Left* and *Right* cortical activity during *Engage*. Unlike Aranyi et al. (2016), our results indicate a lateralization ($Left > Right$ alpha power) during *View*, but no lateralization during *Engage*.

Considering the increase in cortical activity on both *Left* and *Right* sides from *View* to *Engage* (although, not a statistically significant increase on the *Right* side), our results suggest an increase in mental effort from *View* to *Engage*. This is in line with our self-reported data examining the mental demand of both components. Participants used a Likert scale taken from the NASA Task Load Index (NASA TLX) (Hart, 2006), an assessment of mental workload, to rate mental effort necessary to perform both *View* and *Engage* tasks. We conducted a one-tailed paired samples *t*-test [$t_{(18)} = -3.366, p = 0.002$] and determined that *Engage* mental demand ($M = 4.74, SD = 1.59$) was significantly greater than *View* mental demand ($M = 2.63, SD = 1.83$).

4.2. Unsuccessful blocks characterized by comparable increases in *Right* and *Left* prefrontal cortical activity

Of the 142 unsuccessful blocks, 6% were characterized by a decrease in FAA. Therefore, unsuccessful blocks were largely characterized by no change in FAA from *View* to *Engage*. We investigated the contribution of *Left* and *Right* alpha power in unsuccessful blocks as well, and determined a statistically significant interaction between *Side* and *Component Type* [$\Lambda = 0.896, F_{(1,141)} = 16.428, p < 0.001, \eta_p^2 = 0.104$]. Simple main effects analysis of *Side* was also significant [$\Lambda = 0.795, F_{(1,141)} = 36.467, p < 0.001, \eta_p^2 = 0.205$], indicating that *Left* alpha power ($M = 2.00, SD = 1.72$) was significantly lower than *Right* alpha power ($M = 2.68, SD = 2.03$) during *View*. This is the opposite of our results for successful blocks, in which *Left* alpha power was significantly greater than *Right* alpha power during *View*. It appears that unsuccessful blocks may have instead been characterized by a higher starting FAA during *View*,

and therefore, it may have been significantly more difficult to increase FAA further during *Engage*.

Unlike with successful blocks, we additionally determined a statistically significant difference [$\Lambda = 0.873$, $F_{(1,141)} = 20.460$, $p < 0.001$, $\eta_p^2 = 0.127$] between *Left* ($M = 1.72$, $SD = 1.37$) and *Right* alpha power ($M = 2.11$, $SD = 1.57$) during *Engage*, indicative of greater *Left* than *Right* cortical activity during *Engage* for unsuccessful blocks. Our results indicate that the lack of a significant increase in FAA from *View* to *Engage* found in unsuccessful blocks may have been influenced by the greater increase in *Right* than *Left* cortical activity from *View* to *Engage*, as seen in Figure 11.

A simple main effect analysis of *Component Type* was also statistically significant [$\Lambda = 0.769$, $F_{(1,141)} = 42.252$, $p < 0.001$, $\eta_p^2 = 0.231$]. We determined that *Right* alpha power during *View* ($M = 2.68$, $SD = 2.03$) was significantly greater than *Right* alpha power during *Engage* ($M = 2.11$, $SD = 1.57$), indicative of greater increased *Right* cortical activity during *Engage* than during *View*. *Left* alpha power during *View* ($M = 2.00$, $SD = 1.72$) was significantly greater than *Left* alpha power during *Engage* ($M = 1.72$, $SD = 1.37$), indicative of greater *Left* cortical activity during *Engage* than during *View* [$\Lambda = 0.900$, $F_{(1,141)} = 15.736$, $p < 0.001$, $\eta_p^2 = 0.100$]. Therefore, self-reported data concerning mental demand of *View* and *Engage* blocks is in line with unsuccessful blocks as well, in that more cortical processing took place during *Engage* on both *Right* and *Left* sides.

Results suggest that while successful blocks were characterized by a relatively greater increase in *Left* compared to *Right* cortical activity, unsuccessful blocks were characterized by a comparable increase in both *Right* and *Left* cortical activity. While successful blocks may have also been characterized by a relatively lower starting FAA during *View*, making an increase in FAA more feasible during *Engage*, unsuccessful blocks may have been characterized by a higher starting FAA during *View*, perhaps making an increase in FAA during *Engage* more difficult. Considering that there was also greater cortical activity on both sides during *Engage* for unsuccessful blocks, it is possible that participants exerted greater mental effort during unsuccessful blocks as alpha power is thought to be inversely related to cortical network activity (Allen et al., 2004; Smith et al., 2017). Mean *View* and *Engage* FAA values for each block for each participant are shown in Table 1.

4.3. Inter-subject correlations (ISC) suggest varying strategies for achieving NF success

ISCs and EEG forward modeling were calculated to explore spatial distributions of sources of neural activity and to assess the between subject reliability of evoked responses during *View* and *Engage* blocks for each story block. As subject response to visual

feedback is not a stationary process, ISCs were calculated in 1 s windows across all subjects for each story block and decomposed into the three largest correlation components. Percentages of windows which were significant above chance were calculated and reported in Table 2. Components during *View* showed a larger percentage of significant windowed correlation with the exception of story blocks #2, #6, and #7, which showed an equal or larger percentage of significant windows during *Engage*. Results suggest that subject neural activity was significantly less correlated during *Engage*, as neural activity showed longer durations of significant correlated activity during *View*, with the exception of story blocks #2, #6, and #7. Longer de-correlated activity during *Engage* could be due to different thought content strategies for changing the main character's color saturation employed by different subjects. Higher correlations during *View* are not unexpected as subject counting strategies are likely more consistent than strategies taken during the NF task.

Blocks were split into successful and unsuccessful groups containing all successful or unsuccessful blocks respectively. Forward EEG models of ISC of alpha power quantifying largest correlation components across all subjects were calculated for each group. As expected, forward models verify that successful blocks show increased left-frontal activity relative to right-frontal activity compared to unsuccessful blocks (Figure 12). These data suggest that subject success on a given block was marked by robust and reliable asymmetric activity.

4.4. Self-reported data

We collected self-reported data from our questionnaire to inform us of the human experience of brain activity modulation with the experimental BCI story environment. In this section, we present our results concerning thought content strategies for brain activity modulation, perceived success in BCI engagement, and levels of enjoyment and frustration.

4.4.1. Participant strategies

All participants were asked to describe their strategies during *Engage* blocks. We first investigated *Direct* and *Indirect* strategies during *Engage* (Aranyi et al., 2016). We defined *Direct* strategies as those which directly involved interacting with the main character or the story environment. Examples of *Direct* strategies include “protecting [the main character] and making her feel safe,” “speak[ing] with her using my thoughts,” and “creating exit strategies as well as trying to comfort or help the girl escape.” *Indirect* strategies instead involved thoughts that were unrelated to the story environment, such as recalling past experiences and memories. Examples of *Indirect* strategies include thinking of “something that makes me happy, like the color purple... or my friends,” and “[having] a conversation with myself.” We determined that seven participants used *Indirect*

TABLE 1 The mean FAA values of the distribution of 15 FAA values, for each *View* and *Engage* Component of each story block, for each participant.

ID	Story Block #1		Story Block #2		Story Block #3		Story Block #4		Story Block #5		Story Block #6		Story Block #7		Story Block #8	
	View	Engage	View	Engage	View	Engage	View	Engage	View	Engage	View	Engage	View	Engage	View	Engage
#1	0.115	−0.267	−0.041	−0.182	−0.125	0.003	−0.098	0.172	0.188	0.058	−0.031	0.018	0.120	0.198	0.029	−0.077
#2	−0.193	0.057	0.150	0.068	−0.012	0.114	0.092	0.290	0.035	0.202	0.186	0.232	0.058	0.089	0.070	0.017
#3	−0.511	−0.114	−0.626	−0.535	−0.576	−0.595	−0.825	−0.374	−0.560	−0.484	−0.434	−0.563	−0.473	0.067	−0.173	−0.548
#4	0.416	0.484	0.415	0.587	0.812	0.202	0.119	0.464	0.400	0.615	0.508	0.239	0.091	0.291	0.588	0.684
#5	−0.610	−1.040	−0.872	−1.070	−0.944	−1.099	−1.058	−0.847	−0.996	−0.608	−0.815	−0.993	−0.981	−0.561	−0.894	−1.101
#6	0.908	1.408	0.977	1.055	0.853	1.221	1.090	0.697	1.084	0.927	0.911	0.787	0.872	0.924	1.059	1.000
#7	0.237	0.339	0.496	0.326	0.063	0.557	0.353	0.212	0.271	0.251	0.205	0.032	0.349	0.471	0.352	0.095
#8	−0.497	−0.406	−0.020	−0.447	−0.065	−0.478	−0.340	−0.468	−0.134	−0.272	−0.269	−0.273	−0.231	−0.448	−0.216	−0.235
#9	0.493	0.632	0.339	0.408	0.371	0.458	0.644	0.560	0.454	0.473	0.513	0.570	0.592	0.478	0.423	0.629
#10	0.667	0.271	0.690	0.207	0.576	0.408	0.611	0.465	0.485	0.288	0.663	0.504	0.570	0.522	0.687	0.287
#11	−0.105	−0.104	−0.048	0.176	0.206	0.198	0.423	0.089	0.185	0.177	0.623	0.044	−0.096	0.086	0.503	−0.077
#12	0.645	0.396	0.696	0.576	0.593	1.057	0.762	0.482	0.911	0.814	0.387	0.481	0.673	0.742	0.618	0.423
#13	0.038	−0.268	−0.582	−0.087	−0.122	0.257	0.223	0.114	−0.684	0.051	0.064	−0.203	0.053	−0.107	−0.446	−0.163
#14	0.635	0.826	0.475	0.649	0.677	0.840	0.839	0.777	0.561	0.509	0.602	0.810	0.695	0.528	0.573	0.759
#15	1.070	0.701	1.216	0.474	0.950	0.891	1.161	0.852	1.065	0.882	0.858	0.982	0.878	1.179	1.138	0.656
#16	−0.048	0.218	0.248	0.190	0.141	0.105	0.545	0.494	0.116	0.142	−0.132	0.462	0.232	0.427	0.154	0.291
#17	1.076	0.776	0.988	0.632	0.851	0.797	0.595	0.679	0.818	0.515	0.602	0.597	0.704	0.726	0.747	0.630
#18	−0.039	−0.218	0.294	0.176	−0.033	0.442	0.291	0.281	0.428	0.269	0.636	0.236	−0.200	0.059	0.330	−0.006
#19	0.606	0.228	0.700	0.402	0.788	0.332	0.438	0.321	0.105	0.412	0.499	0.251	0.417	0.426	0.599	0.326

TABLE 2 Percentage of windows which show significant intersubject correlation for each View and Engage component of each story block for the three largest correlation component projections.

Component	Percentage Significant ISC Windows											
	Story Block #1		Story Block #2		Story Block #3		Story Block #4		Story Block #5		Story Block #6	
	View	Engage	View	Engage	View	Engage	View	Engage	View	Engage	View	Engage
#1	13.51	10.81	5.41	9.46	14.86	9.46	21.62	16.89	14.19	9.46	6.76	22.97
#2	15.54	8.78	1.35	11.49	22.97	8.78	2.70	.00	10.14	14.19	7.43	10.81
#3	9.46	8.78	4.73	4.05	9.46	6.08	8.78	4.73	6.08	10.14	5.41	9.46

strategies while 12 participants used *Direct* strategies. We used a Chi-Squared test and determined that there was no significant association between block success (Success/Failure) and strategy (*Direct/Indirect*), $X^2(1, N = 19) = 0.833$, $p = 0.361$, Cramer's $V = 0.209$. In future work, it would be necessary to control for strategy type in order to make conclusions about any relationship between strategy and NF interaction success.

Our data shows five out of seven successful participants used *Direct* strategies, while six out of 12 unsuccessful participants used *Direct* strategies. Similarly, Aranyi et al. (2016) determined that *Indirect* strategies were more prominent for unsuccessful participants. Although block success was not significantly associated with strategy type, our RQ3 is partially supported, as a majority of successful participants used *Direct* strategies during *Engage*, while half of unsuccessful participants used *Indirect* strategies.

4.4.2. Perceived success, enjoyment, frustration

Participants also provided subjective ratings of their success during *Engage* blocks using an 8-point Likert scale. We used a Pearson Bivariate one-tailed correlation and determined no statistically significant correlation between perceived success and number of successful blocks $r_{(19)} = -0.089$, $p = 0.358$, suggesting that participants were unclear about NF success. However, enjoyment ($M = 5.89, SD = 1.66$) was significantly greater than frustration ($M = 2.79, SD = 2.28$), as determined by a one-tailed paired t -test [$t_{(18)} = 4.083, p < 0.001$].

5. Conclusion

Our results indicate that seven out of 19 participants were able to use thought strategies to modulate brain activity (FAA) when engaging with the complex story-based BCI environment. Results suggest that successful *Engage* blocks are characterized by both a large increase in FAA from *View* to *Engage*, and a large effect size, while unsuccessful blocks are characterized by relatively no change in FAA from *View* to *Engage*. Considering both successful and unsuccessful blocks, prefrontal cortical activity on left and right sides was significantly greater during *Engage* than *View*, in line with self-reported data which demonstrates participants found *Engage* blocks significantly more mentally demanding than *View* blocks. With our ISC analysis, we determined that components during *View* showed a larger percentage of significant correlation time than components during *Engage*. Therefore, it is likely that patterns of participants' neural activity were more similar during *View*, in which participants counted, than during *Engage*, in which thought content strategies were relatively unconfined. During narrative film viewing, attention has been shown to modulate similar EEG evoked responses (Cohen et al., 2017).

Forward models of largest correlation components between successful and unsuccessful blocks

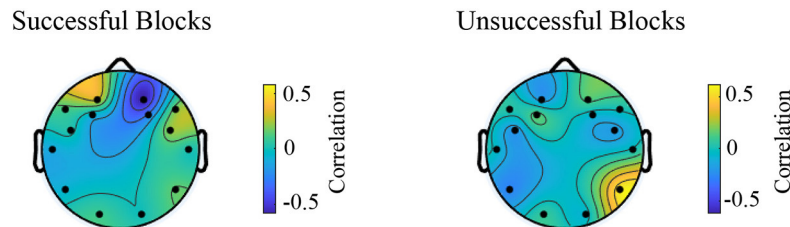


FIGURE 12

Spatial distribution of between subject correlations shows no consistent pattern of neural activity across story blocks in engage and view states.

Therefore, it is possible that participants differed significantly in their ability to pay attention during *Engage* blocks.

Our investigation of left and right prefrontal cortical activities during successful blocks suggests that an increase in left cortical activity, rather than a decrease in right cortical activity from *View* to *Engage* was influential in increased FAA scores from *View* to *Engage*, which suggests successful blocks may have been the result of increased approach motivation. Data suggests increasing right and left prefrontal cortical activity from *View* to *Engage* contributed to the unchanging FAA score for unsuccessful blocks. Forward models of decomposed electrical activity of strongest correlated responses also confirmed that successful blocks were marked with robust left-right asymmetries. It is possible that blocks in the unsuccessful category have some level of asymmetry, but at lower levels than successful blocks. As such, future work in FAA based BCIs should titrate the level of left-right asymmetric activity with reliable FAA detection and controllability.

While our results indicate multiple successful blocks across ~37% of participants, the percentage of unsuccessful participants (63%) in this study is higher than typical BCI illiteracy rates at 15–30% (Saha et al., 2021). However, BCI illiteracy, when participants cannot modulate brain activity within the time frame of the study, is not well-understood. We offer several ideas concerning the difference in successful blocks between our study and that of Aranyi et al. (2016). First, our consumer-grade EEG device was likely not sensitive enough to detect small changes in FAA, as indicated by successful blocks characterized by moderate to large effect sizes only. Second, our self-reported results concerning perceived success indicate that participants did not understand when they were successful in modulating brain activity, which may arise as a diminished neural response ISC seen in *Engage* compared to *View* data. Because success was unclear, participants may not have felt rewarded by visual feedback of their brain activity. Reward is essential for learning through operant conditioning,

and may explain some of the lack of BCI engagement success.

Considering work by Hasson et al. (2008), it is possible that certain film works may elicit more similar, homogeneous brain responses across viewers, as seen through higher ISCs during certain films over others. In their study, the authors determined that higher ISCs were found when participants viewed a film directed by an accomplished film director, compared to unstructured video recorded in a park environment. Hasson et al. (2008) conclude that films which can better guide attention may lead to greater ISCs. Additionally, it is known that films played backwards do not have as high ISCs as those played in their original temporal order (forward, typically; Hasson et al., 2008), or when scenes are scrambled in time (Dmochowski et al., 2012). Because the experimental story presented was designed to be abstract, unguided, and open-ended to interpretation, our limited ISCs results are expected. Additionally, it is likely that the open-ended nature of the experimental narrative contributed to the comparable ISCs during *View* and *Engage* blocks as well, as the unguided nature of the story blocks led to higher variances in brain activity than might have been found with story blocks with a more concrete, guided story. While a more guided story BCI may have elicited greater ISC, future research is needed to determine how ISC relates to NF interaction success with different kinds of BCI experiences.

Previous research has explored participants' prior BCI experience, as it "could possibly influence [the participant's] participation in the study" (Zioga et al., 2018; p. 4). Considering that none of the participants had previously interacted with a BCI, it is possible that participants expected a different, perhaps mentally passive experience, in which brain activity was recorded "as is," without the effort necessary to engage with NF, such as the BCI experience of Zioga et al. (2018). Although our goal was to provide participants with open-ended ideas for NF interaction, story blocks may not have been salient enough for participants to achieve high levels of NF success. While some

participants expanded on the experimental story narrative in creative ways through self-reported data, others were unable to find any meaning or narrative structure within what one participant referred to as “moving images.” Questionnaire data suggests that participants did not apply one strategy consistently throughout *Engage* blocks, as they did in the case of Aranyi et al. (2016). Additional research concerning strategy consistency in BCI engagement would elucidate our findings. Similar to Aranyi et al. (2016), we did not control for valence (positive or negative affect) of NF strategies. Questionnaire data showed that participants used sad as well as happy thought content strategies during *Engage*, which likely influenced NF success rates. According to Friedrich et al. (2014) however, participants can be divided into three groups: one third of participants who gain control of brain signals immediately, one third who cannot gain control of brain signals, and *one third who can only gain control of brain signals after training*. With this in mind, it is possible that our system did not allow for adequate training for individuals who needed training to be successful, as only one third of the participants achieved NF interaction success in our study. Research has shown that NF success can be driven by differences in learner (participant) strategies. For example, Davelaar et al. (2018) found that learners (participants who could gain control of their brain activity) described feeling, or being aware of something, while non-learners (participants who did not gain control of their brain activity during the course of the training session) exerted significant mental effort or tried to maintain strict attentional focus. Davelaar et al. (2018) describe these strategies on a feeling-sensing continuum, where NF success increases when moving from trying to sensing. Therefore, our provided instructions, however open-ended, may have prompted participants to try numerous different mental activities, whereas simply “feeling” could have been more effective.

However, results indicate that it is possible for participants to engage with an affective BCI by increasing FAA while experiencing a complex, experimental story environment with imagination-based instructions. It is possible that only those who need less training can be successful with NF interaction with the current system. Despite lower NF success rates than that of Aranyi et al. (2016), participants indicated higher enjoyment than frustration in self-reported data. The role of strategy type (direct or indirect) in FAA NF success is not clear in the present study, and should be investigated in future work. Our study contributes to the goal of making BCI experiences more user-friendly, enjoyable and motivating, while investigating experimental storytelling techniques with which to engage participants during affective BCI interaction.

5.1. Limitations and future directions

Our study is limited in a number of different ways. First, we did not collect self-reported data concerning

participants' current mood, anxiety, or depression symptoms. Our sample was also very homogeneous: undergraduate students from our department. Ethnic data was not collected. One participant reported currently undergoing psychiatric treatment. Additionally, participants were not asked to refrain from caffeine, cigarettes, alcohol or drugs within 12 h of the experiment. Similar to work by Gapen et al. (2016), this study was intended to be ecologically valid, therefore, its exclusion criteria were limited. However, a positive, effortlessly relaxed mental state may be best for NF interaction in BCIs (Friedrich et al., 2014), therefore, negative moods may have contributed to lack of success seen with our system. Collecting self-reported mood ratings before and after the NF protocol (Zotef et al., 2020) would additionally allow us to learn more about the participant experience. Considering that the participant's motivation, locus of control as well as empathy levels can play a role in BCI performance (Friedrich et al., 2014), a greater focus on participant characteristics would inform both NF success rates with experimental BCIs, and design techniques for future BCI experiences.

In future work, we would like to explore BCI usage in an environment outside the university or departmental setting, with the goal of investigating a more diverse group of participants. A replication of this study using a medical-grade EEG device may also allow us to learn more about participants' success rate during NF interaction, although likely within a less comfortable, quick, and convenient experience and time frame for participants.

Although designed to stimulate the imagination, many factors of the experimental story BCI may have counteracted NF success for certain participants. Although not mentioned as confusing or disorienting to participants, it is possible that the manipulation of spatial logic throughout story blocks contributed to a lower NF success rate through a general sense of displacement. Complex stimuli, such as the developed multi-component experimental story interface, may have effects on the learner (participants) which are not yet understood (Enriquez-Geppert et al., 2017). A follow up study could examine experimental storytelling components individually, in order to understand the influence of each component. A few examples of storytelling components to investigate include story block order, depth of spatio-temporal discontinuity, distance between two characters, vividness of environmental objects, and realism of 3d models among a plethora of other possibilities.

Individuals clearly differ in their preference for story genre and style. In future work, we believe an investigation of learner (participant) preferences for story ambiguity will be central in understanding the influence of experimental storytelling in BCI experiences. Future work could also explore guiding participant attention more directly within stories of varying ambiguity. Lastly, examining NF success with blocks of longer duration could allow participants who need more training to be successful. Instead of 30 s, a longer NF block duration, such as 60 s (Dehghani et al., 2020) may

be necessary for some participants to increase FAA within the block.

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: https://osf.io/uqgp8/?view_only=f6df8369f68c47c59ea0f21308ddba50.

Ethics statement

The studies involving human participants were reviewed and approved by Purdue University Institutional Review Board. 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

CK developed the concept, Unity environment, and conducted data collection. CK and BC performed EEG processing and analyses and wrote the manuscript. CK, CM, and BC completed statistical analyses and revised the manuscript.

CM acquired funding for the study. All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

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.

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.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2022.883467/full#supplementary-material>

References

- Albu, C. (2020). Intimate connections: Alternative communication threads in nina sobell's video performances and installations (1974–82). *Camera Obscura* 35, 39–75. doi: 10.1215/02705346-8085111
- Allen, J. J., Coan, J. A., and Nazarian, M. (2004). Issues and assumptions on the road from raw signals to metrics of frontal EEG asymmetry in emotion. *Biol. Psychol.* 67, 183–218. doi: 10.1016/j.biopsycho.2004.03.007
- 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
- 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: ACM), 511–521. doi: 10.1145/2807442.2807447
- Aranyi, G., Cavazza, M., and Charles, F. (2015a). "Using FNIRs for prefrontal-asymmetry neurofeedback: methods and challenges," in *International Workshop on Symbiotic Interaction* (Berlin: Springer), 7–20. doi: 10.1007/978-3-319-24917-9_2
- Ben-Israel, A., and Greville, T. N. (2003). *Generalized Inverses*. New York, NY: Springer-Verlag.
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Berger, A. M., and Davelaar, E. J. (2018). Frontal alpha oscillations and attentional control: a virtual reality neurofeedback study. *Neuroscience* 378, 189–197. doi: 10.1016/j.neuroscience.2017.06.007
- Briesemeister, B. B., Tamm, S., Heine, A., and Jacobs, A. M. (2013). Approach the good, withdraw from the bad—a review on frontal alpha asymmetry measures in applied psychological research. *Psychology* 4, 261–267. doi: 10.4236/psych.2013.43A039
- Cattan, G. (2021). The use of brain-computer interfaces in games is not ready for the general public. *Front. Comput. Sci.* 3:628773. doi: 10.3389/fcomp.2021.628773
- 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. (2014a). "Towards empathic neurofeedback for interactive storytelling," in *2014 Workshop on Computational Models of Narrative* (Quebec City, QC: Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik).
- Cavazza, M., Charles, F., Aranyi, G., Porteous, J., Gilroy, S. W., Raz, G., et al. (2014b). "Towards emotional regulation through neurofeedback," in *Proceedings of the 5th Augmented Human International Conference* (Kobe), 1–8. doi: 10.1145/2582051.2582093
- Charles, F., Martins, C. D. C., and Cavazza, M. (2020). Prefrontal asymmetry bci neurofeedback datasets. *Front. Neurosci.* 14:601402. doi: 10.3389/fnins.2020.601402
- 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
- Cohen, S. S., Henin, S., and Parra, L. C. (2017). Engaging narratives evoke similar neural activity and lead to similar time perception. *Sci. Rep.* 7, 1–10. doi: 10.1038/s41598-017-04402-4
- Davelaar, E. J., Barnby, J. M., Almasi, S., and Eatough, V. (2018). Differential subjective experiences in learners and non-learners in frontal alpha neurofeedback: piloting a mixed-method approach. *Front. Hum. Neurosci.* 2018:402. doi: 10.3389/fnhum.2018.00402

- David, O. A., Predatu, R., and Maffei, A. (2021). Ret Hink online video game for children and adolescents: effects on state anxiety and frontal alpha asymmetry. *Int. J. Cogn. Ther.* 14, 399–416. doi: 10.1007/s41811-020-00077-4
- Dehghani, A., Soltanian-Zadeh, H., and Hossein-Zadeh, G.-A. (2020). Global data-driven analysis of brain connectivity during emotion regulation by electroencephalography neurofeedback. *Brain Connect.* 10, 302–315. doi: 10.1089/brain.2019.0734
- 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
- Dmochowski, J. P., Sajda, P., Dias, J., and Parra, L. C. (2012). Correlated components of ongoing EEG point to emotionally laden attention—a possible marker of engagement? *Front. Hum. Neurosci.* 6:112. doi: 10.3389/fnhum.2012.00112
- 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
- Fischer, R. A. (1971). *The Design of Experiments*, 9th Edn. New York, NY: Macmillan Publishing Co.
- Friedrich, E. V., Wood, G., Scherer, R., and Neuper, C. (2014). Mind over brain, brain over mind: cognitive causes and consequences of controlling brain activity. *Front. Hum. Neurosci.* 8:348. doi: 10.3389/fnhum.2014.00348
- Gapen, M., van der Kolk, B. A., Hamlin, E., Hirshberg, L., Suvak, M., and Spinazzola, J. (2016). A pilot study of neurofeedback for chronic PTSD. *Appl. Psychophysiol. Biofeedb.* 41, 251–261. doi: 10.1007/s10484-015-9326-5
- Gilroy, S. W., Porteous, J., Charles, F., Cavazza, M., Soreq, E., Raz, G., et al. (2013). “A brain-computer interface to a plan-based narrative,” in *Twenty-Third International Joint Conference on Artificial Intelligence* (Beijing).
- Gonzalez-Franco, M., Egan, Z., Peachey, M., Antley, A., Randhava, T., Panda, P., et al. (2020). “Movebox: democratizing MOCAP for the microsoft rocketbox avatar library,” in *2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)* (Utrecht: IEEE), 91–98. doi: 10.1109/AIVR50618.2020.00026
- Gruzelić, J., Inoue, A., Smart, R., Steed, A., and Steffert, T. (2010). Acting performance and flow state enhanced with sensory-motor rhythm neurofeedback comparing ecologically valid immersive VR and training screen scenarios. *Neurosci. Lett.* 480, 112–116. doi: 10.1016/j.neulet.2010.06.019
- Harmon-Jones, E., and Gable, P. A. (2018). On the role of asymmetric frontal cortical activity in approach and withdrawal motivation: an updated review of the evidence. *Psychophysiology* 55:e12879. doi: 10.1111/psyp.12879
- Hart, S. G. (2006). “NASA-task load index (NASA-TLX); 20 years later,” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 50 (Los Angeles, CA: Sage publications), 904–908. doi: 10.1177/154193120605000909
- Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., and Heeger, D. J. (2008). Neurocinematics: the neuroscience of film. *Projections* 2, 1–26. doi: 10.3167/proj.2008.020102
- Jensen, C. B. F., Petersen, M. K., Larsen, J. E., Stopczynski, A., Stahlhut, C., Ivanova, M. G., et al. (2013). “Spatio temporal media components for neurofeedback,” in *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)* (San Jose, CA: IEEE), 1–6. doi: 10.1109/ICMEW.2013.6618362
- Johnston, S., Linden, D. E. J., Healy, D., Goebel, R., Habes, I., and Boehm, S. (2011). Upregulation of emotion areas through neurofeedback with a focus on positive mood. *Cogn. Affect. Behav. Neurosci.* 11, 44–51. doi: 10.3758/s13415-010-0010-1
- Kelley, N. J., Hortensius, R., Schutter, D. J., and Harmon-Jones, E. (2017). The relationship of approach/avoidance motivation and asymmetric frontal cortical activity: a review of studies manipulating frontal asymmetry. *Int. J. Psychophysiol.* 119, 19–30. doi: 10.1016/j.ijpsycho.2017.03.001
- Kerous, B., Skola, F., and Liarokapis, F. (2018). EEG-based BCI and video games: a progress report. *Virt. Real.* 22, 119–135. doi: 10.1007/s10055-017-0328-x
- Kisker, J., Lange, L., Flinkenflügel, K., Kaup, M., Labersweiler, N., Tetenborg, F., et al. (2021). Authentic fear responses in virtual reality: a mobile EEG study on affective, behavioral and electrophysiological correlates of fear. *Front. Virt. Real.* 2:716318. doi: 10.3389/frvir.2021.716318
- Koenitz, H. (2017). “Beyond ‘walking simulators’-games as the narrative avant-garde,” in *Digital Games Research Association Conference* (Melbourne, VIC).
- Kosmyna, N., and Lécuyer, A. (2017). Designing guiding systems for brain-computer interfaces. *Front. Hum. Neurosci.* 11:396. doi: 10.3389/fnhum.2017.00396
- Kuper, N., Käckemester, W., Wacker, J., and Fajkowska, M. (2019). Resting frontal EEG asymmetry and personality traits: a meta-analysis. *Eur. J. Pers.* 33, 154–175. doi: 10.1002/per.2197
- Lackner, N., Unterrainer, H.-F., Skirris, D., Shaheen, S., Dunitz-Scheer, M., Wood, G., et al. (2016). EEG neurofeedback effects in the treatment of adolescent anorexia nervosa. *Eating Disord.* 24, 354–374. doi: 10.1080/10640266.2016.1160705
- Le Groux, S., Manzolli, J., Verschure, P. F., Sanchez, M., Luvizotto, A., Mura, A., et al. (2010). “Disembodied and collaborative musical interaction in the multimodal brain orchestra,” in *NIME* (Sydney, NSW), 309–314.
- Le, T. P., Lucas, H. D., Schwartz, E. K., Mitchell, K. R., and Cohen, A. S. (2020). Frontal alpha asymmetry in schizotypy: electrophysiological evidence for motivational dysfunction. *Cogn. Neuropsychiatry* 25, 371–386. doi: 10.1080/13546805.2020.1813096
- Mennella, R., Patron, E., and Palomba, D. (2017). Frontal alpha asymmetry neurofeedback for the reduction of negative affect and anxiety. *Behav. Res. Ther.* 92, 32–40. doi: 10.1016/j.brat.2017.02.002
- Mori. (2003). *Wave UFO*, Vol. 3. Available online at: <https://www.publicartfund.org/exhibitions/view/wave-ufo/> (accessed June 10, 2022).
- Nijholt, A. (2015). “Competing and collaborating brains: multi-brain computer interfacing,” in *Brain-Computer Interfaces*, eds A. E. Hassanien and A. T. Azar (New York, NY: Springer), 313–335. doi: 10.1007/978-3-319-10978-7_12
- Parra, L. C., Spence, C. D., Gerson, A. D., and Sajda, P. (2005). Recipes for the linear analysis of EEG. *NeuroImage* 28, 326–341. doi: 10.1016/j.neuroimage.2005.05.032
- Peeters, F., Ronner, J., Bodar, L., van Os, J., and Lousberg, R. (2014). Validation of a neurofeedback paradigm: manipulating frontal EEG alpha-activity and its impact on mood. *Int. J. Psychophysiol.* 93, 116–120. doi: 10.1016/j.ijpsycho.2013.06.010
- Pike, M., Wilson, M. L., Benford, S., and Ramchurn, R. (2016). “# scanners: a BCI enhanced cinematic experience,” in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, CA), 293–296. doi: 10.1145/2851581.2889468
- Quaedflieg, C. W., Smulders, F. T., Meyer, T., Peeters, F., Merckelbach, H., and Smeets, T. (2016). The validity of individual frontal alpha asymmetry EEG neurofeedback. *Soc. Cogn. Affect. Neurosci.* 11, 33–43. doi: 10.1093/scan/nsv090
- Robinson, R., Wiley, K., Rezaeivahdati, A., Klarkowski, M., and Mandryk, R. L. (2020). “Let’s get physiological, physiological!” A systematic review of affective gaming,” in *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 132–147. doi: 10.1145/3410404.3414227
- 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. Self-Regul.* 20, 241–258. doi: 10.1007/BF01474516
- Saha, S., Mamun, K. A., Ahmed, K. I. U., Mostafa, R., Naik, G. R., Darvishi, S., et al. (2021). Progress in brain computer interface: challenges and potentials. *Front. Syst. Neurosci.* 15:4. doi: 10.3389/fnsys.2021.578875
- Sitaram, R., Ros, T., Stoelckel, 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
- Smith, E. E., Reznik, S. J., Stewart, J. L., and Allen, J. J. (2017). Assessing and conceptualizing frontal EEG asymmetry: an updated primer on recording, processing, analyzing, and interpreting frontal alpha asymmetry. *Int. J. Psychophysiol.* 111, 98–114. doi: 10.1016/j.ijpsycho.2016.11.005
- Szczelkun, S. (2018). *SENSE THINK ACT: A Collection Of Exercises to Experience Total Human Ability*, Vol. 1. London: Routine Art Co.
- Taberham, P. (2018). *Lessons in Perception: The Avant-Garde Filmmaker as Practical Psychologist*. New York, NY: Berghahn Books. doi: 10.2307/j.ctv3znzvc
- Templeton, G. F. (2011). A two-step approach for transforming continuous variables to normal: implications and recommendations for is research. *Commun. Assoc. Inform. Syst.* 28:4. doi: 10.17705/1CAIS.02804
- Wadson, A., Nijholt, A., and Nam, C. S. (2015). Artistic brain-computer interfaces: state-of-the-art control mechanisms. *Brain Comput. Interfaces* 2, 70–75. doi: 10.1080/2326263X.2015.1103155
- Zioga, P., Pollick, F., Ma, M., Chapman, P., and Stefanov, K. (2018). ‘Enheduanna a manifesto of falling’ live brain-computer cinema performance: performer and audience participation, cognition and emotional engagement using multi-brain BCI interaction. *Front. Neurosci.* 12:191. doi: 10.3389/fnins.2018.00191
- Zotef, V., Mayeli, A., Misaki, M., and Bodurka, J. (2020). Emotion self-regulation training in major depressive disorder using simultaneous real-time fMRI and EEG neurofeedback. *NeuroImage* 27:102331. doi: 10.1016/j.neuroimage.2020.102331



OPEN ACCESS

EDITED BY

Chang-Hwan Im,
Hanyang University, South Korea

REVIEWED BY

Fangzhou Xu,
Qilu University of Technology, China
Jane Zhen Liang,
Shenzhen University, China

*CORRESPONDENCE

Marco Congedo
marco.congedo@gipsa-lab.fr

SPECIALTY SECTION

This article was submitted to
Brain-Computer Interfaces,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 21 September 2022

ACCEPTED 11 November 2022

PUBLISHED 02 December 2022

CITATION

Bleuzé A, Mattout J and Congedo M
(2022) Tangent space alignment:
Transfer learning for Brain-Computer
Interface.
Front. Hum. Neurosci. 16:1049985.
doi: 10.3389/fnhum.2022.1049985

COPYRIGHT

© 2022 Bleuzé, Mattout and Congedo.
This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License](#)
(CC BY). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Tangent space alignment: Transfer learning for Brain-Computer Interface

Alexandre Bleuzé¹, Jérémie Mattout² and Marco Congedo^{1*}

¹GIPSA-Lab, University Grenoble Alpes, CNRS, Grenoble INP, Grenoble, France, ²Lyon Neuroscience Research Center, INSERM, CNRS, University Claude Bernard Lyon 1, Lyon, France

Statistical variability of electroencephalography (EEG) between subjects and between sessions is a common problem faced in the field of Brain-Computer Interface (BCI). Such variability prevents the usage of pre-trained machine learning models and requires the use of a calibration for every new session. This paper presents a new transfer learning (TL) method that deals with this variability. This method aims to reduce calibration time and even improve accuracy of BCI systems by aligning EEG data from one subject to the other in the tangent space of the positive definite matrices Riemannian manifold. We tested the method on 18 BCI databases comprising a total of 349 subjects pertaining to three BCI paradigms, namely, event related potentials (ERP), motor imagery (MI), and steady state visually evoked potentials (SSVEP). We employ a support vector classifier for feature classification. The results demonstrate a significant improvement of classification accuracy, as compared to a classical training-test pipeline, in the case of the ERP paradigm, whereas for both the MI and SSVEP paradigm no deterioration of performance is observed. A global 2.7% accuracy improvement is obtained compared to a previously published Riemannian method, Riemannian Procrustes Analysis (RPA). Interestingly, tangent space alignment has an intrinsic ability to deal with transfer learning for sets of data that have different number of channels, naturally applying to inter-dataset transfer learning.

KEYWORDS

Brain-Computer Interface, Riemannian geometry, transfer learning, domain adaptation, ERP, motor imagery, SSVEP

1. Introduction

A Brain-Computer Interface (BCI) is a system that allows interactions between a human and a machine using only neurophysiological signals coming from the brain. It aims at rehabilitating, improving, or enhancing the ability of the user by means of a computerized system (Wolpaw et al., 2002). The most common modality used to record neurophysiological signals is electroencephalography (EEG). This is mainly because EEG is affordable, completely safe for the user and because it features a high temporal resolution. EEG signals can be translated into a command to be sent to a computer by means of a decoding algorithm. The loop is often closed by means of a feedback given to the user.

Several BCI applications have emerged to help patients, such as spellers (Yin et al., 2015; Rezeika et al., 2018) or wheelchair controllers (Li et al., 2013). The focus in this line of research is to restore lost communication or movement capabilities. Other applications are designed for the rehabilitation of patients after an incapacitating event such as a stroke (Frisoli et al., 2012; Mane et al., 2020). Non-clinical applications have also been proposed, for example to provide a means of control in video games (Congedo et al., 2011; Bonnet et al., 2013). Mixed approaches are also possible, for example in the Cybathlon BCI Race, where people with complete or severe loss of motor function compete in a video game-based competition (Perdikis et al., 2018).

Several paradigms can be used in order to control a BCI. The most commons are event related potentials (ERP), motor imagery (MI), and steady state visually evoked potentials (SSVEP). The ERP paradigm consists of electrical potentials evoked by sensory stimulations; in MI the user imagines to move body parts, resulting in synchronizations and desynchronizations in the sensory-motor cortex; in SSVEP-based BCIs, the user concentrates on visual stimuli flashed at distinct frequencies, leading to responses at the same frequency in the brain. Regardless of the paradigm, it is necessary to calibrate the BCI system in order to allow proper decoding. The calibration process is time consuming, annoying for the healthy user and problematic for the clinical population, which has limited mental resources (Mayaud et al., 2016). In fact, a calibration is required not only for every new user, but also for every new session of the same user. This is due to the high inter-subject and inter-session variability of the features extracted from the EEG. Such variability is caused by several factors, including, but not limited to, the impedance and placement of EEG electrodes, individual morphological and physiological characteristics of the brain and changing brain states.

One way to deal with this variability is to use transfer learning (TL). This means trying to reuse some of the information we have already gathered on known data that may be coming from either previous subjects or previous sessions. In transfer learning we usually consider two types of data. The **source** represents the data we already know on a given subject whereas the **target** consists of a new subject whose some training data may be available, but mostly is unlabeled and is to be used as a test. The aim is to adapt as accurately as possible the data of the target using the few available training data to the source data (or vice versa). In order to do so, several methods have already been developed.

The authors in Jayaram et al. (2016) adapt the weights given to spatial features that are meant to predict the stimulus in order to transfer information from one subject to another or from one session to another. Some other methods adapt the parameters of a neural network. For example the authors in Fahimi et al. (2019) perform a partial retraining of a deep neural network on a small number of samples of a new user, improving significantly the accuracy. Unsupervised domain adaptation methods have

also emerged, as in Sun et al. (2015), where the authors perform unsupervised transfer learning in the Euclidean domain, using covariance matrices to align data from different subjects. A well-established approach for classification in the BCI field is to use covariance matrices of the signal since those matrices have many relevant properties (Congedo et al., 2017). The covariance matrices are Symmetric Positive Definite (SPD) and therefore lie in a Riemannian manifold. In this way, some algorithms have been developed to achieve transfer learning in the Riemannian manifold of SPD matrices. For instance, the authors in Zanini et al. (2018) propose a recentering procedure consisting in translating the center of mass of both the source and target data to the identity using parallel transport. This procedure is actually equivalent to a whitening using the Riemannian mean as anchor point. In Yair et al. (2019), both the center of mass of the source and target data are translated to their midpoint along the geodesic, allowing equivalent results. The authors of Rodrigues et al. (2019), inspired by the Procrustes analysis, proposed to add two more steps after recentering: a stretching of the observations, so as to equalize the dispersion of the data in the source and target domain and a rotation, so as to align as much as possible the center of mass of each class between the source and the target data set. The method, named Riemannian Procrustes Analysis (RPA), was shown to allow efficient transfer learning. A later alignment method was discussed in He and Wu (2020). This method is similar to Sun et al. (2015) with improvement related to enhanced dimensions in the Euclidean space. The authors of Zhang et al. (2020) chose another approach by transferring instances of the source close enough to the target in order to enhance the low data availability of the target model. They used MI data and compared the proximity of source and target trials using Hamming distance after preprocessing steps. Another idea proposed in Zhang and Wu (2020) is to find a common subspace between source and target, yielding a projection matrix to reduce the gap between the source and the target. Finally the authors train on the source subspace to test on target subspace.

In this article we introduce a Riemannian transfer learning approach similar in spirit to the RPA approach (Rodrigues et al., 2019), but operating in the tangent space. Our contribution has multiple benefits as compared to previous attempts. First, it lies in a state-of-the-art BCI feature space, the Riemannian tangent space, introduced in the BCI domain by Barachant et al. (2010). Since the tangent space is an Euclidean space, there exists a wide variety of well-established tools to decode the data therein and in general they are faster as compared to a decoding approach in the Riemannian manifold. Second, since it acts on an Euclidean space, it can be used for all kind of feature vectors, not just those obtained in a Riemannian setting. Third, our method is computationally effective, as it only requires one singular value decomposition (SVD). Fourth, it extends naturally to the heterogeneous transfer learning case, i.e., when the number and/or placements of electrodes is not the same in the source

and target data set. In a similar previous attempt the SVD has been applied independently on the source and target dataset and the resulting matrices are then used to align the data (Sun et al., 2015). Our method is instead casted as a Procrustes problem and therefore it fulfills a well-known optimality condition for inter-domain alignment.

A previous version of our method has been presented in Bleuzé et al. (2021). As compared to that presentation, we have improved it by adding several ways to deal with the rank deficiency of the cross product matrix. Also, here we test it on a very large amount of data, namely, 18 BCI databases comprising a total of 349 subjects. Furthermore, these databases pertain to three BCI modalities: event related potentials (ERP), motor imagery (MI), and steady state visually evoked potentials (SSVEP). Therefore, the present study is a comprehensive test bed, which is a well-grounded way to reach general conclusions when comparing machine learning pipelines.

2. Materials and methods

2.1. Notations

Throughout this article we will denote matrices with upper case bold characters (\mathbf{A}), vectors with lower case bold characters (\mathbf{a}), indices and scalars by lower case italic characters (a), and constants by upper case italics (A). The function $\text{tr}(\cdot)$ will indicate the trace of a matrix, $(\cdot)^T$ its transpose, $\|\cdot\|$ the 2-norm or the Frobenius norm, \circ the Hadamard product, $\log(\cdot)$, and $\exp(\cdot)$ the matrix logarithm and exponential, respectively. \mathbf{I}_N will denote the identity matrix in dimension N .

2.2. Riemannian geometry

Let us consider a set of trials $\{\mathbf{X}_n\}_{n \in [1, N]}$ with shape (N_c, N_{sample}) , where N_c is the number of channels, N_{sample} the number of (temporal) samples and N the number of matrices in the set. A generic trial is simply denoted as \mathbf{X} . In order to be as close as possible to a realistic scenario, we consider data with a low level of pre-processing and we do not use any artifact removal method, such as ocular artifacts or outliers removal (Çinar and Acir, 2017; Minguillon et al., 2017).

The (spatial) sample covariance matrix estimation (SCM) writes

$$\mathbf{C} = \frac{1}{N_{\text{sample}} - 1} \mathbf{X} \mathbf{X}^T. \quad (1)$$

The SCM has shape (N_c, N_c) . It lies in a Riemannian manifold of symmetric positive definite (SPD) matrices (Bhatia, 2009). It is therefore possible to classify directly a set $\{\mathbf{C}_n\}$ of covariance matrices by means of classification algorithms acting on such a manifold, such as the Minimum Distance to Riemannian Mean (MDRM) classifier (Barachant et al., 2012) or its refinement

Riemannian Minimum Distance to Means Field (RMDMF; Congedo et al., 2019). It is also possible to project the matrices onto the tangent space of the manifold at a base point \mathbf{M} and use Euclidean classifiers therein (Barachant et al., 2012, 2013). The base point \mathbf{M} in this work will always be chosen as the Log-Euclidean mean, which is defined as Fillard et al. (2005).

$$\mathbf{M} = \exp \left(\frac{1}{N} \sum_n \log(\mathbf{C}_n) \right). \quad (2)$$

The projection onto the tangent space at base point \mathbf{M} is obtained by the logarithmic map operator (Nielsen and Bhatia, 2013).

$$\text{Log}_{\mathbf{M}}(\mathbf{C}) = \mathbf{M}^{\frac{1}{2}} \log \left(\mathbf{M}^{-\frac{1}{2}} \mathbf{C} \mathbf{M}^{-\frac{1}{2}} \right) \mathbf{M}^{\frac{1}{2}}. \quad (3)$$

The projected matrix is now a (N_c, N_c) symmetric matrix. Since we are concerned with transfer learning (TL), we are interested in matching the position of the source and target data sets in the manifold as much as possible. Following Zanini et al. (2018), we recenter both the source and target data sets by setting their global mean at the identity. This is simply obtained by transforming all trials of a dataset such as

$$\mathbf{C}_{\text{rec}} = \mathbf{M}^{-\frac{1}{2}} \mathbf{C} \mathbf{M}^{-\frac{1}{2}}, \quad (4)$$

where \mathbf{M} is the center of mass of the observations and \mathbf{C}_{rec} denotes the recentered trial. After recentering all trials their center of mass becomes the identity matrix, corresponding to the “zero” point in an Euclidean space.

The logarithmic mapping at the identity simplifies, yielding

$$\begin{aligned} \text{Log}_{\mathbf{I}_{N_c}}(\mathbf{C}_{\text{rec}}) &= \mathbf{I}_{N_c}^{1/2} \log \left(\mathbf{I}_{N_c}^{-1/2} \mathbf{M}^{-\frac{1}{2}} \mathbf{C} \mathbf{M}^{-\frac{1}{2}} \mathbf{I}_{N_c}^{-1/2} \right) \mathbf{I}_{N_c}^{1/2} \\ &= \log \left(\mathbf{M}^{-\frac{1}{2}} \mathbf{C} \mathbf{M}^{-\frac{1}{2}} \right). \end{aligned}$$

The above recentering followed by tangent space projection was first proposed in the BCI field in Barachant et al. (2012, 2013) and is nowadays a standard processing procedure, which in this article is carried out systematically, unless explicitly mentioned.

Once projected onto the tangent space the matrices are vectorized. Since they are symmetric, only the upper (or lower) triangle of the matrix is kept and the off-diagonal terms are weighted by $\sqrt{2}$ so as to preserve the norm of the original matrix. In mathematical notation, the vectorization of tangent vector \mathbf{S} reads

$$\mathbf{s} = \text{triu}(\mathbf{S} \circ \mathbf{A}), \quad (5)$$

with $\text{triu}(\cdot)$ the operator vectorizing the upper triangle and \mathbf{A} a matrix with the same shape as \mathbf{S} , filled with 1 on the diagonal and $\sqrt{2}$ on the off-diagonal part. Since the matrices have been previously recentered, the resulting vectors are also recentered, that is, the mean tangent vector is the zero vector.

Having obtained the tangent vectors as described here above, it is possible to use all the well know classification algorithms

that act in an Euclidean space, the most commonly employed in the BCI community being the linear discriminant analysis (LDA; Barachant et al., 2012), support vector classifier (SVC; Xu et al., 2021) and Lasso logistic regression (LR) (Tomioka et al., 2006). In this study, we use the SVC.

2.3. Alignment

As anticipated in the introduction, the method here proposed has been inspired by previous Riemannian TL methods, such as in Zanini et al. (2018) and Rodrigues et al. (2019) which focus on covariance matrices in the Riemannian space of SPD matrices. In Rodrigues et al. (2019) the authors consider again the recentring and add two further alignment steps:

1. a *rescaling* so as to match the dispersion around the mean (center of mass) in both the source and target data sets and
2. a *rotation* so as to align the mean of each class as much as possible. This effectively results in a Riemannian Procrustes alignment.

In this article, the same steps are undertaken in the tangent space. In particular, we rotate the tangent vectors using an Euclidean Procrustes procedure.

Let us consider the set of centered tangent vectors for the source $\{\mathbf{s}_n\}_{n \in [1, N_s]}$ and the target $\{\mathbf{t}_n\}_{n \in [1, N_t]}$ domain. N_s and N_t are the number of vectors for, respectively, the source and the target data set. As we will see, in the following it will not be required that the source and target tangent vectors have the same dimensions. Denoting \mathbf{s} and \mathbf{t} the generic source and target tangent vectors, the rescaling is obtained setting the average norm within each set to 1, which is readily obtained by transformations

$$\tilde{\mathbf{s}} = \frac{\mathbf{s}}{\frac{1}{N_s} \sum_n \|\mathbf{s}_n\|} \quad (6)$$

and

$$\tilde{\mathbf{t}} = \frac{\mathbf{t}}{\frac{1}{N_t} \sum_n \|\mathbf{t}_n\|}, \quad (7)$$

yielding rescaled source and target data sets $\{\tilde{\mathbf{s}}_n\}_{n \in [1, N_s]}$ and $\{\tilde{\mathbf{t}}_n\}_{n \in [1, N_t]}$. It is also possible to set the norm of the target data set equal to the norm of the source data set if it is sought not to modify the norm of the source data set.

For the rotation (alignment), we propose a supervised method that uses the mean point of the classes. Let us consider K classes that we ought to align. Although other procedures are possible, in the following we always align the target set to the source set. Let \mathbf{y} and \mathbf{z} be the label vectors of, respectively, $\{\tilde{\mathbf{s}}_n\}_{n \in [1, N_s]}$ and $\{\tilde{\mathbf{t}}_n\}_{n \in [1, N_t]}$ with shape N_s and N_t . We start by

computing the mean for each class k , given its N_k trials

$$\bar{\mathbf{s}}_k = \frac{1}{N_k} \sum_{\mathbf{y}_i=k} \tilde{\mathbf{s}}_i \quad (8)$$

for the source set and

$$\bar{\mathbf{t}}_k = \frac{1}{N_k} \sum_{\mathbf{z}_i=k} \tilde{\mathbf{t}}_i \quad (9)$$

for the target set. In the supervised procedure these vectors are the *anchor points* we use for alignment. Therefore, we define

$$\bar{\mathbf{S}} = [\bar{\mathbf{s}}_k, k \in [1, K]] \quad (10)$$

and

$$\bar{\mathbf{T}} = [\bar{\mathbf{t}}_k, k \in [1, K]] \quad (11)$$

as the two matrices of shape $\left(\frac{N_c(N_c+1)}{2}, K\right)$ holding the anchor vectors stacked one next to the other. We can now define the cross-product matrix

$$\mathbf{C}_{\text{st}} = \bar{\mathbf{S}} \bar{\mathbf{T}}^T. \quad (12)$$

of shape $\left(\frac{N_c(N_c+1)}{2}, \frac{N_c(N_c+1)}{2}\right)$. Like any rectangular matrix—or squared when source and target have the same number of channels— \mathbf{C}_{st} can be decomposed by singular value decomposition, such that

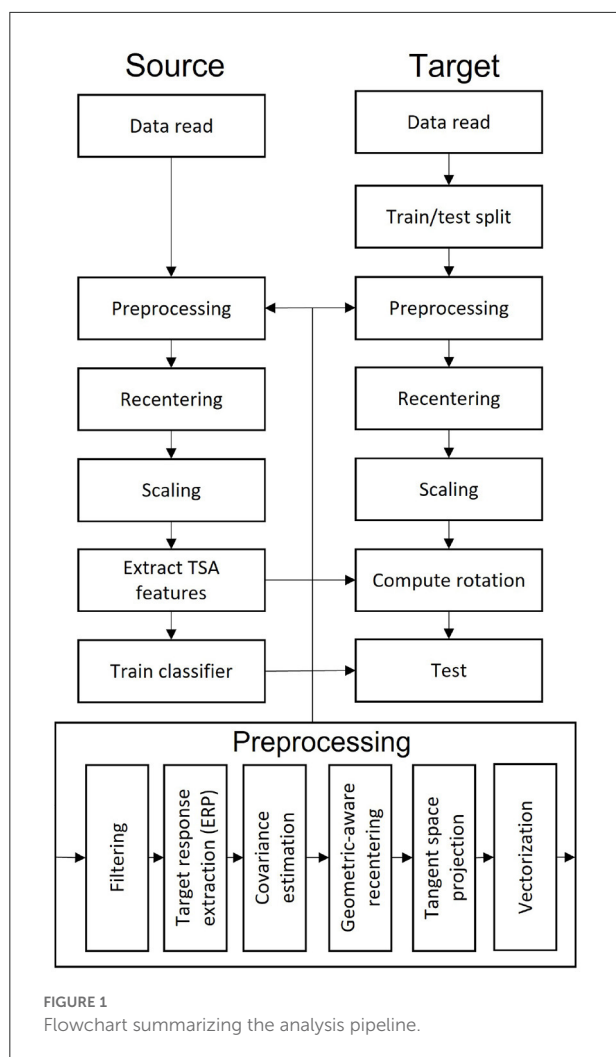
$$\mathbf{C}_{\text{st}} = \mathbf{U} \mathbf{D} \mathbf{V}^T, \quad (13)$$

with \mathbf{U} and \mathbf{V} the two orthonormal matrices holding in columns the left and right singular vectors, respectively, and \mathbf{D} a diagonal matrix holding the singular values. As usual in signal processing, we will retain a subset of the singular vectors in order to suppress noise. Such a truncation has also the advantage to work for the case where \mathbf{U} and \mathbf{V} do not have the same shape. As a general rule, we seek the smallest number N_v of singular vectors which corresponding singular values explain at least 99.9% of the variance, resulting in $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{V}}$ with shape $\left(\frac{N_c(N_c+1)}{2}, N_v\right)$. Finally, we are able to align the target vectors previously created $\{\tilde{\mathbf{t}}_n\}_{n \in [1, N]}$ to the domain of the source vectors $\{\tilde{\mathbf{s}}_n\}_{n \in [1, N]}$ as

$$\hat{\mathbf{t}} = \tilde{\mathbf{U}} \tilde{\mathbf{V}}^T \tilde{\mathbf{t}} \quad (14)$$

where $\hat{\mathbf{t}}$ denotes the aligned target vectors. The newly created set $\{\hat{\mathbf{t}}_n\}_{n \in [1, N]}$ is now aligned to the space of source vectors $\{\tilde{\mathbf{s}}_n\}_{n \in [1, N]}$, therefore it can be classified with algorithms trained on the source domain. As it is well-known, when the cross-product in Equation (12) is full-rank, the unique solution to the Procrustes optimization problem

$$\arg \min_{\mathbf{Z}} (\|\mathbf{Z} \bar{\mathbf{T}} - \bar{\mathbf{S}}\|) \quad (15)$$

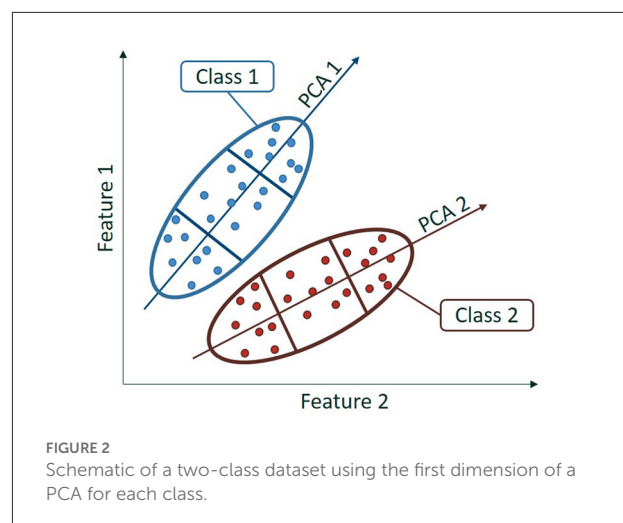


is indeed $\mathbf{Z} = \mathbf{UV}^T$. In our case, the solution is not unique. Note that a connection between this problem and the Bures-Wasserstein metric has been recently described in [Bhatia et al. \(2019\)](#).

The whole process for our method is summarized in [Figure 1](#).

2.4. Augmenting/improving \mathbf{C}_{st}

Cross-product matrix \mathbf{C}_{st} is usually rank deficient and its estimation could be improved in several ways. In this section, we will suggest two such improvements. First, as long as a supervised TL is possible, since we are relying on averaging the tangent vectors, it is possible to employ robust average estimators. For instance, we may consider the trimmed means,



the median or power means to estimate suitable anchor points. We may also stack several of these average estimators to obtain larger matrices $\bar{\mathbf{s}}_k$ and $\bar{\mathbf{t}}_k$, which may provide more robust information on the actual central tendency of the data.

Second, we cluster the data sets in several subsets describing the shape of the data set when considered altogether and compute separate means for each cluster. We may use, for instance, principal component analysis (PCA) on each class independently to create clusters, as depicted in [Figure 2](#). The centroid of those clusters are then computed and used as anchor points. In order to obtain the clustering for both the source and target data set, we consider for each class a PCA trained on the source and used as such on both the source and target data set.

Using such a clustering procedure, if the source and the target data set display a rather similar shape, their alignment will be very effective, leading to promising transfer learning results. Such a procedure is also possible with unlabeled data in case of unsupervised TL. However, in this case, we have noticed that at least two PCA components are necessary to obtain an efficient transfer learning. Therefore, in the unsupervised case we recommend using at least two PCA components and separate data for each dimensions, creating a \mathbf{C}_{st} matrix with shape $\left[\frac{N_c(N_c+1)}{2}, N_d \times N_g \right]$ with N_d the number of dimensions used and N_g the number of groups created in each dimensions. An effective strategy is to visualize the data and their representations in order to verify whether the chosen reduced dimensionality offers a good approximation of the data as it may be as well totally inaccurate, depending on the data, especially for unsupervised TL. For our results, we chose to create three PCA clusters for each class and use these means to compute the cross-product matrix \mathbf{C}_{st} as it gives enough information without reducing the size of the data used for each mean too much.

TABLE 1 List of the processed databases for event-related potentials, motor imagery paradigms, and steady-state visual evoked potentials.

	Type	Ch	Trials	Sess	Sub
2003–2015	P300	8	1,800	1	10
2008–2014	P300	8	4,200	1	8
2009–2014	P300	16	1,728	3	10
Brain Invaders 2013a	P300	16	480	1	24
EPFL P300 dataset	P300	32	3,268	4	8
2001–2014	MI	22	288	2	9
2001–2015	MI	13	400	2	12
2002–2014	MI	15	160	1	14
2004–2014	MI	3	720	5	9
Alexandre motor imagery	MI	16	40	1	8
Cho 2017	MI	64	200	1	49
Grosse-Wentrup 2009	MI	128	300	1	10
Physionet motor imagery	MI	64	45	1	109
Weibo 2014	MI	60	160	1	10
Zhou 2016	MI	14	290	3	4
SSVEP exoskeleton	SSVEP	8	160	1	12
SSVEP Nakanishi	SSVEP	8	180	1	9
SSVEP Wang	SSVEP	62	240	1	34

Ch, number of channels; Sess, number of sessions; Sub, number of subjects.

3. Results

The TSA algorithm previously introduced has been tested on three well know BCI paradigms: ERP, MI, and SSVEP. We have analyzed 18 open-access BCI databases available on the Mother Of All BCI Benchmark (MOABB; Jayaram and Barachant, 2018). Python library, of which five uses the ERP paradigm, 10 the MI paradigm, and 3 the SSVEP paradigm. The 18 databases include a total of 349 subjects with very high variability between and within datasets. We summarize the data in Table 1, following Congedo et al. (2019).

We execute transfer learning from one subject to the other for all possible source/target pair of subjects within each database. The accuracy is evaluated using balanced accuracy since the number of trials per class is often unbalanced (always for the ERP paradigm).

Since the amount of data of all pair-wise comparisons is huge, we start by visually evaluating all balanced accuracies obtained for a given database by means of seriation plots, i.e., plots showing the accuracy for each pair of target and source subject arranged in matrix form. The accuracy is averaged over all numbers of alignment trials for each pair. The case where the target and the source are the same, i.e., on the diagonal, is replaced by the classical train-test cross-validation accuracy, offering a straightforward benchmark. Furthermore, the target and source subjects are sorted on rows by descending order of the train-test cross-validation accuracy. It should be kept

in mind that since the train-test procedure is fully supervised and optimized for the train data, it is expected to outperform a transfer learning method. Figure 3 shows a representative seriation plot for each paradigm allowing the visual comparison of the performance of the TSA vs. the RPA transfer learning method; all figures are available as Supplementary material.

In order to evaluate the average performance, we plot the balanced accuracy averaged across all subjects in a database for each method as a function of the number of alignment trials. Since we are averaging across subjects, for this analysis we include only those subjects featuring at least 60% accuracy in a classical train-test cross-validation. This restriction excludes about half of the subjects, leaving 178 subject out of 349. It's important to note that subjects with an overall 60% accuracy usually have more than 70% accuracy with all available training trials. Figure 4 shows a representative plot for each paradigm. All figures are available as Supplementary material.

Then, we summarize all the pair-wise accuracy information in the accuracy tables such as Table 2. These tables give for each target subject the accuracy averaged over numbers of alignment trials and source subjects. For this database, there are 30 numbers of alignment trials considered and 18 possible source subjects, which makes an average over 540 values. This makes the standard error low in general. Accuracy tables for all databases are given as Supplementary material.

The accuracy tables confirm what can be evaluated visually in the average accuracy plots and seriation plots; on the average there is about 1% difference between classical train-test cross-validation accuracy and TSA and about 5% between classical train-test cross-validation and RPA. This speaks in favor of a clear improvement of the TSA method over the RPA method. Table 3 summarizes all the balanced accuracy for each dataset. On the average across databases there is no loss of accuracy using a TSA as compared to the optimal train-test accuracy. This is not true for the RPA.

Finally, we performed statistical tests on all pair-wise source/target accuracy results we have collected. To this end, we follow the procedure introduced in Rodrigues et al. (2021). In a nutshell: we first compute signed paired *t*-test for every target subject comparing the accuracy between methods, yielding *T* statistics $T_{m,i}$ and *p*-values $p_{m,i}$ for each pair of methods *m* and target subject *i*. In order to correct for the multiplicity of statistical tests we use Holm's sequential rejection multiple test procedure (Holm, 1979) for each target subject. This produces tables such as Table 4. Corresponding tables for each database are available as Supplementary material. Then, we combine the *p*-values we obtain using the Stouffer's Z-score method (Zaykin, 2011) for each database, yielding multiple *p*-values corresponding to each pair of methods for each database. Those *p*-values are also corrected by means of Holm's procedure and are summarized in Table 5. In this table we can see that among the 18 databases we have analyzed, 11 show a significant improvement of TSA as compared to RPA.

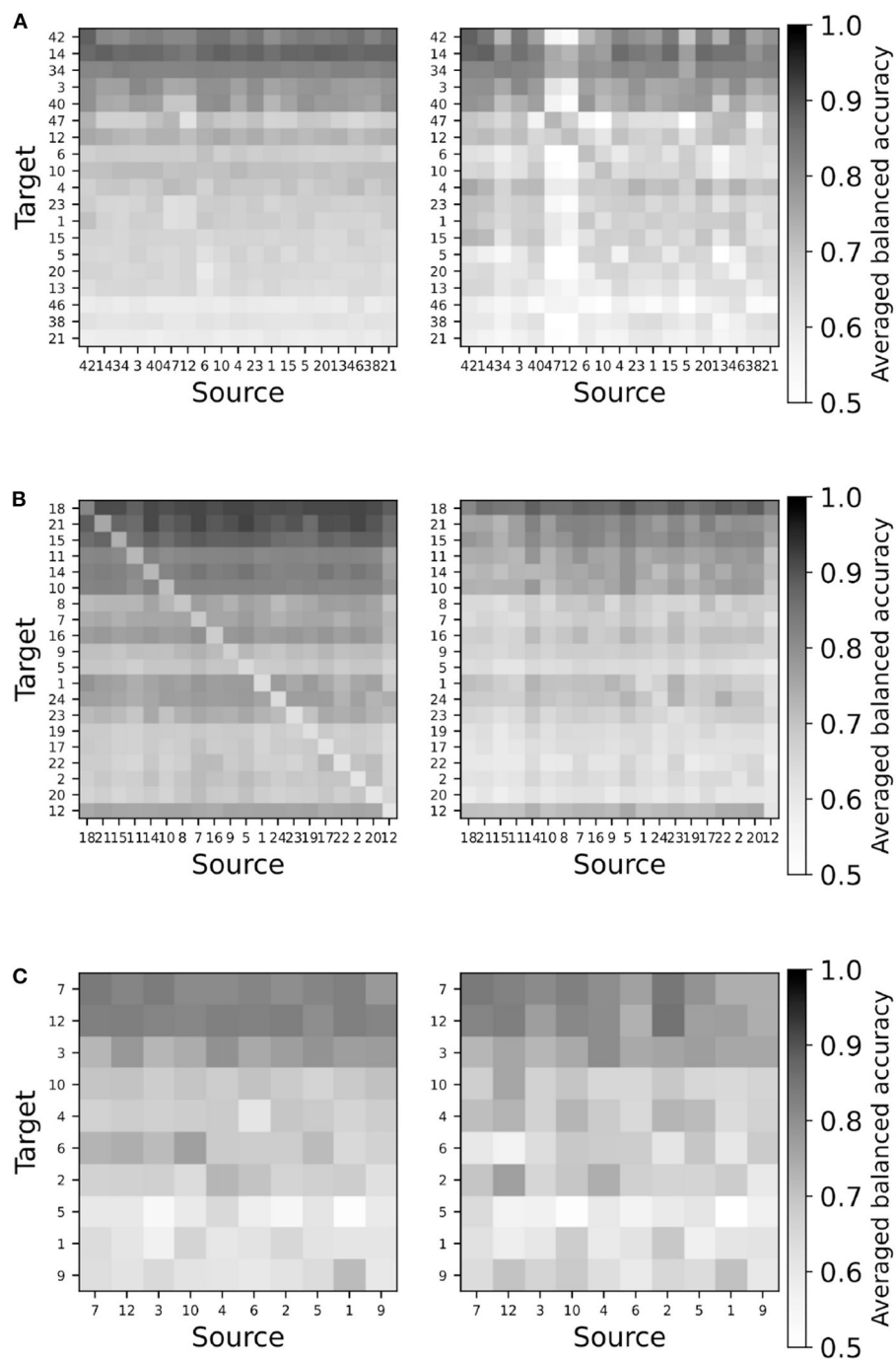


FIGURE 3

Representative seriation plots for TSA (left) and RPA (right) methods for each paradigms. See text for details. (A) Database Cho2017 (MI). (B) Database brain invaders 2013a (ERP). (C) Database SSVEP exoskeleton (SSVEP).

No significant difference between the classical train-test cross-validation accuracy and TSA is found with the exception of two databases, for which TSA proves inferior. This number

grows to five databases comparing the classical train-test cross-validation accuracy and the accuracy obtained by RPA. Finally, using Stouffer's Z-score method, p -values corresponding to each

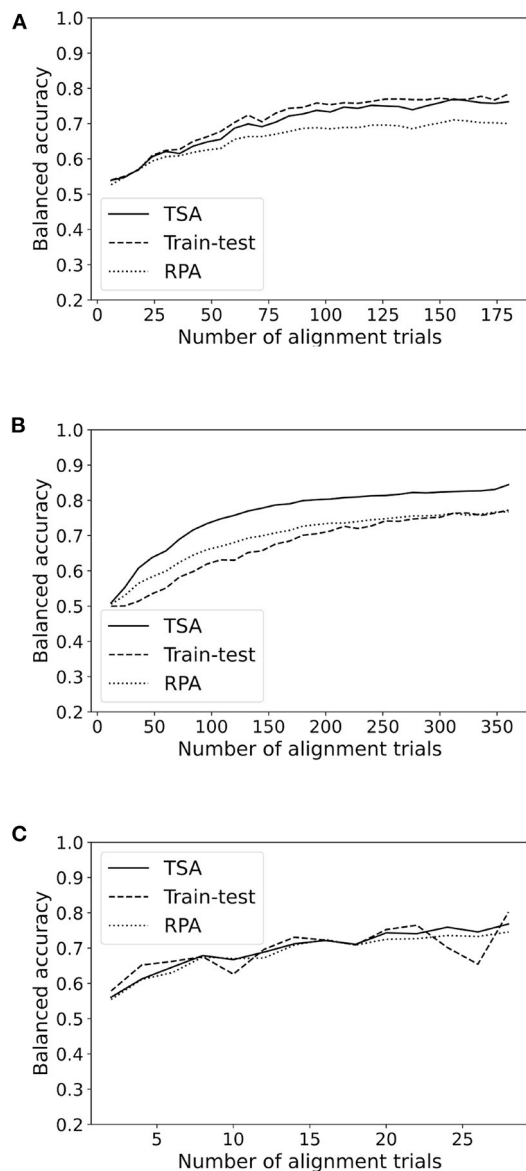


FIGURE 4
Accuracy as a function of the number of alignment trials for TSA and RPA methods for a representative database in each paradigm (MI, ERP, SSVEP). **(A)** Database Cho2017 (MI). **(B)** Database brain invaders 2013a (ERP). **(C)** Database SSVEP exoskeleton (SSVEP).

paradigm are computed and corrected with Holm's procedure (Table 6).

So far we have focused on cross-subject transfer learning, however the method we propose can also be used to transfer different information, such as from one session to another with the same subject or from one task to the another. In order to ensure the ability of our method to reduce the inter-session variability, we used the dataset having multiple sessions and

TABLE 2 Balanced accuracy \pm standard error for the good subjects of database Cho 2017 averaged over number of alignment trials and source for train-test, TSA, and RPA.

	Train-test acc	TSA acc	RPA acc
Subject 1	67.86 \pm 2.00	66.70 \pm 0.43	65.84 \pm 0.43
Subject 3	81.39 \pm 0.81	77.78 \pm 0.18	74.96 \pm 0.32
Subject 4	69.26 \pm 1.49	69.24 \pm 0.40	69.25 \pm 0.50
Subject 5	65.90 \pm 2.12	65.03 \pm 0.50	59.12 \pm 0.55
Subject 6	70.74 \pm 2.11	67.65 \pm 0.38	59.80 \pm 0.53
Subject 10	70.55 \pm 1.91	70.71 \pm 0.45	62.32 \pm 0.43
Subject 12	70.97 \pm 1.61	73.01 \pm 0.39	67.68 \pm 0.29
Subject 13	63.60 \pm 1.14	64.30 \pm 0.25	60.83 \pm 0.25
Subject 14	88.26 \pm 1.52	86.99 \pm 0.37	82.06 \pm 0.45
Subject 15	66.15 \pm 1.64	65.63 \pm 0.38	65.20 \pm 0.37
Subject 20	64.96 \pm 1.13	64.64 \pm 0.26	60.79 \pm 0.33
Subject 21	60.63 \pm 1.50	58.63 \pm 0.24	56.54 \pm 0.30
Subject 23	68.64 \pm 1.62	66.69 \pm 0.37	64.94 \pm 0.41
Subject 34	82.73 \pm 0.37	81.89 \pm 0.09	79.97 \pm 0.15
Subject 38	61.98 \pm 1.27	61.51 \pm 0.27	58.53 \pm 0.35
Subject 40	76.73 \pm 1.89	76.30 \pm 0.44	72.21 \pm 0.47
Subject 42	88.34 \pm 3.32	83.02 \pm 0.73	75.78 \pm 0.77
Subject 46	62.55 \pm 1.79	59.08 \pm 0.37	54.64 \pm 0.26
Subject 47	72.34 \pm 1.66	67.98 \pm 0.36	63.03 \pm 0.42

TABLE 3 Balanced accuracy \pm standard error for the good subjects of each dataset averaged over number of alignment trials, target and source for train-test, TSA, and RPA.

	Train-test acc	TSA acc	RPA acc
2003–2015	69.04 \pm 0.74	76.05 \pm 0.57	69.63 \pm 0.57
2008–2014	67.52 \pm 0.69	72.00 \pm 0.38	65.50 \pm 0.38
2009–2014	70.48 \pm 0.50	77.38 \pm 0.21	70.90 \pm 0.21
Brain invaders 2013a	67.26 \pm 0.45	76.07 \pm 0.11	69.59 \pm 0.11
EPFL P300 dataset	67.22 \pm 0.61	73.013 \pm 0.24	65.28 \pm 0.24
2001–2015	81.13 \pm 0.73	79.01 \pm 0.24	76.21 \pm 0.24
2002–2014	78.40 \pm 0.70	76.51 \pm 0.20	72.55 \pm 0.55
2004–2014	74.35 \pm 1.30	76.00 \pm 0.59	75.27 \pm 0.59
Alexandre motor imagery	78.51 \pm 2.82	76.30 \pm 1.33	75.93 \pm 1.33
Cho 2017	71.24 \pm 0.52	69.83 \pm 0.12	65.97 \pm 0.12
Grosse-Wentrup 2009	77.20 \pm 0.91	73.12 \pm 0.34	72.16 \pm 0.34
Physionet motor imagery	70.75 \pm 1.35	64.28 \pm 0.27	63.26 \pm 0.27
Weibo 2014	73.97 \pm 1.25	69.66 \pm 0.50	71.27 \pm 0.50
Zhou 2016	82.82 \pm 1.59	82.43 \pm 0.86	81.31 \pm 0.86
SSVEP exoskeleton	69.50 \pm 1.16	69.69 \pm 0.40	68.72 \pm 0.40
SSVEP Nakanishi	95.55 \pm 0.82	95.95 \pm 0.28	96.83 \pm 0.28
SSVEP Wang	68.26 \pm 2.63	57.42 \pm 0.76	59.11 \pm 0.76
Global	74.42 \pm 0.30	74.49 \pm 0.12	71.79 \pm 0.12

processed results for inter-session cross-validation. In order to do so, we used the data of one session as a source, then for each other sessions with 80% test and 20% training data split we trained the transfer learning model and tested the results. The

TABLE 4 Subject-wise *p*-values for MI database Cho 2017.

	TSA>RPA	TRAIN>TSA	TRAIN>RPA
Subject 1	0.002*	0.818	0.420
Subject 3	0.033*	<0.001*	<0.001*
Subject 4	0.005*	0.690	0.248
Subject 5	<0.001*	0.980	0.301
Subject 6	0.001*	0.379	0.013*
Subject 10	<0.001*	0.503	0.007*
Subject 12	<0.001	0.086	<0.001*
Subject 13	0.001*	0.997	0.700
Subject 14	0.443	<0.001*	<0.001*
Subject 15	0.004*	0.964	0.843
Subject 20	0.385	0.991	0.988
Subject 21	<0.001*	1.000	0.999
Subject 23	0.009*	0.836	0.514
Subject 34	<0.001*	<0.001*	<0.001*
Subject 38	<0.001*	1.000	0.980
Subject 40	0.020*	0.002*	<0.001*
Subject 42	0.361	<0.001*	<0.001*
Subject 46	0.126	1.000	0.995
Subject 47	<0.001*	0.038*	<0.001*

TRAIN, train-test; *Significant *p*-values after multiple comparison correction.

TABLE 5 Database-wise *p*-values.

	TSA>RPA	TRAIN>TSA	TRAIN>RPA
2003–2015	< 0.001	1.000	0.663
2008–2014	0.002*	0.877	0.326
2009–2014	<0.001*	1.000	0.561
Brain invaders 2013a	<0.001*	1.000	1.000
EPFL P300 dataset	<0.001*	1.000	0.102
2001–2014	0.004*	0.468	0.191
2001–2015	<0.001*	0.133	0.005*
2002–2014	<0.001*	0.139	<0.001*
2004–2014	0.515	0.641	0.552
Alexandre motor imagery	0.636	0.353	0.280
Cho 2017	<0.001*	0.098	<0.001*
Grosse-Wentrup 2009	0.097	0.083	0.071
Physionet motor imagery	<0.001*	<0.001*	<0.001*
Weibo 2014	1.000	0.099	0.189
Zhou 2016	0.071	0.453	0.398
SSVEP exoskeleton	0.002*	0.702	0.477
SSVEP Nakanishi	1.000	0.464	0.886
SSVEP Wang	0.989	<0.001*	<0.001*

TRAIN, train-test; *Significant *p*-values after multiple comparison correction.

processed is then repeated using each session as the source. We compared four different methods:

- Tangent Space Alignment (TSA), our method,
- Riemannian Procrustes Analysis (RPA) used as a comparison in this article,

TABLE 6 *p*-values for each paradigm and in global for all tests that have been done. The global *p*-values are the combination of the *p*-values for all databases regardless of their paradigm.

	TSA>RPA	TRAIN>TSA	TRAIN>RPA
p300	<0.001*	1.000	0.961
Imagery	<0.001*	<0.001*	<0.001*
SSVEP	<0.654	<0.001*	<0.001*
Global	<0.001*	1.000	<0.001*

TRAIN, train-test; *Significant *p*-values after multiple comparison correction.

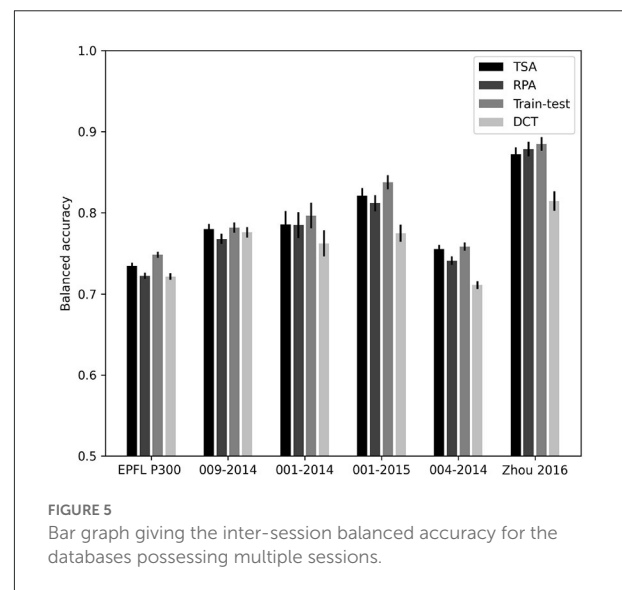


FIGURE 5 Bar graph giving the inter-session balanced accuracy for the databases possessing multiple sessions.

TABLE 7 Inter-session balanced accuracy table for each dataset and methods.

	Train-test	TSA	RPA	DCT
EPFL P300	74.85 ± 0.38	73.46 ± 0.38	72.24 ± 0.39	72.16 ± 0.40
2009–2014	78.19 ± 0.64	78.00 ± 0.65	76.79 ± 0.63	77.59 ± 0.64
2001–2014	79.66 ± 1.57	78.56 ± 1.66	78.48 ± 1.61	76.23 ± 1.63
2001–2015	83.78 ± 0.86	82.10 ± 0.95	81.19 ± 0.99	77.49 ± 1.03
2004–2014	75.83 ± 0.52	75.55 ± 0.50	74.11 ± 0.53	71.09 ± 0.50
Zhou 2016	88.48 ± 0.86	87.22 ± 0.84	87.83 ± 0.90	81.45 ± 1.19

- Usual train-test method using only target data,
- Direct testing (DCT) using algorithms trained on the source without aligning target data by a rotation (recentering only).

The results are given in Figure 5 and presented numerically in Table 7.

These results are coherent with the previous ones. They yield accuracies slightly improved as compared to sessions mixed all together, which is expected. Moreover, the proposed method

performs better than RPA on all databases with the exception of Zhou 2016.

4. Discussion

The extensive analysis we have carried out shows that for the ERP paradigm TSA clearly outperforms RPA. For the MI paradigm we observe that TSA performs better than RPA and that the classical train-test cross-validation outperforms both TSA and RPA. For SSVEP, the classical train-test cross-validation outperforms both TSA and RPA. In global, RPA is outperformed by both TSA and the classical train-test cross-validation. Based on Table 6, we do not conclude on the superiority of TSA as compared to train-test in terms of accuracy as the global value that can be found is mainly due to very large values found for the ERP paradigm. To conclude with the results, it has been shown that TSA outperforms RPA for the large majority of the databases we have used in this analysis, reaching an accuracy pretty close to the optimal train-test method. It is also a clear improvement from a methodological point of view in comparison to RPA as it naturally allows transfer learning between datasets with different number of channels, where RPA needs some extensions in order to do so (Rodrigues et al., 2021).

In this paper, we have introduced a new method for transfer learning inter and intra subject for brain computer interfaces. Our study indicates that it outperforms a state-of-the-art analogous method (RPA). However, it still does not reach the same accuracy that can be achieved with a classical test-train cross-validation procedure for the motor imagery and SSVEP paradigm. Further research is needed to understand why the performance of the TSA method is clearly superior for the ERP paradigm.

In terms of computation time, since we have a closed form for the rotation of TSA method, it is way faster than the RPA method, where an optimization on the Grassmann manifold is performed. However, even if our method can be faster by one order of magnitude, with N_c being the size of the pre-covariance signal (number of channels plus number of channels of the average target for ERP, number of channels for MI, number of classes times number of channels for SSVEP) we compute rotations with size $\left[\frac{N_c*(N_c+1)}{2}, \frac{N_c*(N_c+1)}{2}\right]$ where RPA computes rotation with size (N_c, N_c) . This means that for datasets with a significant amount of channels the computational advantage of TSA will tend to vanish. Table 8 shows the average time of computation for each dataset if no spatial filter were to be applied before the covariance estimation, sometimes yielding big covariance matrices. Usually spatial filter are applied so the low time values are those that are to be usually encountered. Moreover, when computing the rotations for TSA, we can consider only a number of channels that will result in a proportion of

TABLE 8 Database-wise average computation time for both TSA and RPA methods.

	TSA (s)	RPA (s)	Ratio	Cov shape
2003–2015	1.027	11.889	11.57	16 * 16
2008–2014	1.711	10.874	6.36	16 * 16
2009–2014	0.897	6.118	6.82	32 * 32
Brain invaders 2013a	0.774	13.368	17.26	32 * 32
EPFL P300 dataset	1.341	8.607	6.42	64 * 64
2001–2014	0.999	5.075	5.08	22 * 22
2001–2015	0.664	4.362	6.57	13 * 13
2002–2014	0.416	2.241	5.39	15 * 15
2004–2014	0.029	0.166	5.74	3 * 3
Alexandre motor imagery	0.067	0.777	11.61	16 * 16
Cho 2017	5.299	16.858	3.18	64 * 64
Grosse-Wentrup 2009	161.942	113.926	0.70	128 * 128
Physionet motor imagery	2.739	4.432	1.62	64 * 64
Weibo 2014	3.719	13.282	3.57	60 * 60
Zhou 2016	0.468	3.752	8.02	14 * 14
SSVEP exoskeleton	0.056	1.184	21.22	16 * 16
SSVEP Nakanishi	0.41	0.970	23.81	16 * 16
SSVEP Wang	125.949	22.320	0.18	124 * 124

the global variance of the data. This means that we do not need to compute the whole $\left[\frac{N_c*(N_c+1)}{2}, \frac{N_c*(N_c+1)}{2}\right]$ rotation matrix, we can select only a few singular values and their corresponding vectors, highly reducing computation time if needed.

We also observed that in some cases, mainly for ERPs, skipping the rescaling of the target data will lead to improved classification results. Further research is needed to fully understand the role of rescaling in Procrustes-like transfer learning methods.

In this article, we have proposed a procedure to improve the rank deficiency of the cross-product matrix, making the result more stable and more accurate. However, this improvement also presents some downsides. When using only the average point of the data for both source and target data, the method will be pretty sensitive to noise since the number of points used for means computation is drastically reduced. Adding trimmed means and/or medians could make the method less sensitive to noise. Additionally, when augmenting C_{st} using PCA, the more groups will be created, the more each group will be sensitive to noise. By reducing the number of points in a group one increases the impact of artifacts on the average point. Furthermore, PCA is an algorithm that is pretty sensitive to noise and could be replaced by one of its robust variant. It is to be noticed that a low number of groups in general allows a good approximation of the shape of the data. Of course, artifact correction or removal would allow better performances.

It should be reminded that in this article we did not apply any form of pre-processing in order to increase the overall accuracy. We have done so because the goal of the article is to compare transfer learning methods on a large amount of data and in a wide variety of real-world situations, that is, on noisy data. For the same reason, we are using all the data of a subject without making any adaptation from one session to the other. This choice obviously leads to reduced overall accuracy, resulting in an important decrease in the number of subjects used for the final average results. It is important to do so, however, since, as it has been found in previous studies, there is a large gap between “good” and “bad” subjects in transfer learning accuracy (Zanini et al., 2018; Rodrigues et al., 2019).

Another point to mention is that even though our method has been tested extensively on many databases, there are even more databases to test on. Additionally, some paradigms such as affective BCI have not been investigated in this article. Investigation on cross-database transfer learning is still to be done.

As all efficient transfer learning methods, TSA can be very helpful when used along with a machine learning model that takes too much time for training on new data for online sessions. TSA also allows the alignment of multiple subjects' data into the same feature space. Such alignment could improve classification accuracy of multiple subjects and allow the training of robust classifier on aligned data that will give improved results for new subjects once they are aligned. This is the object of current investigation in our laboratory.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: <http://moabb.neurotechx.com/docs/datasets.html>.

References

- Barachant, A., Bonnet, S., Congedo, M., and Jutten, C. (2012). Multiclass brain-computer interface classification by Riemannian geometry. *IEEE Trans. Biomed. Eng.* 59, 920–928. doi: 10.1109/TBME.2011.2172210
- Barachant, A., Bonnet, S., Congedo, M., and Jutten, C. (2013). Classification of covariance matrices using a Riemannian-based kernel for BCI applications. *Neurocomputing* 2013, 172–178. doi: 10.1016/j.neucom.2012.12.039
- Barachant, A., Bonnet, S., Congedo, M., Jutten, C., Riemannian, C. J., and Bonnet, S. (2010). Riemannian geometry applied to BCI classification. *Lecture Notes Comput. Sci.* 6365, 629–636. doi: 10.1007/978-3-642-15995-4_78
- Bhatia, R. (2009). *Positive Definite Matrices*. Princeton, NJ: Princeton University Press. doi: 10.1515/9781400827787
- Bhatia, R., Jain, T., and Lim, Y. (2019). On the Bures-Wasserstein distance between positive definite matrices. *Exposit. Math.* 37, 165–191. doi: 10.1016/j.exmath.2018.01.002
- Bleuze, A., Mattout, J., and Congedo, M. (2021). “Transfer learning for the Riemannian tangent space: applications to brain-computer interfaces,” in *7th International Conference on Engineering and Emerging Technologies, ICEET 2021 (Istanbul)*, 1–6. doi: 10.1109/ICEET53442.2021.9659607
- Bonnet, L., Lotte, F., and Lécuyer, A. (2013). Two brains, one game: design and evaluation of a multiuser BCI video game based on motor imagery. *IEEE Trans. Comput. Intell. AI Games* 5, 185–198. doi: 10.1109/TCIAIG.2012.2237173
- Çınar, S. and Acır, N. (2017). A novel system for automatic removal of ocular artefacts in EEG by using outlier detection methods and independent component analysis. *Expert Syst. Appl.* 68, 36–44. doi: 10.1016/j.eswa.2016.10.009
- Congedo, M., Barachant, A., and Bhatia, R. (2017). Riemannian geometry for eeg-based brain-computer interfaces: a primer and a review. *Brain Comput. Interfaces* 4, 155–174. doi: 10.1080/2326263X.2017.1297192
- Congedo, M., Goyat, M., Tarrin, N., Ionescu, G., Varnet, L., Rivet, B., et al. (2011). “Brain invaders”: a prototype of an open-source p300-based video

Author contributions

AB performed the research, analyzed the data, and wrote the manuscript. MC suggested the original idea for the method. MC and JM designed, reviewed, and edited the manuscript. All authors have read and approved the submitted manuscript.

Funding

This research has been partially supported by CNRS grant 80|PRIMES TrAp and by ANR grant Hifi (ANR-20-CE17-0023).

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.

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.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2022.1049985/full#supplementary-material>

game working with the openvibe platform," in *5th International Brain-Computer Interface Conference, BCI 2011* (Graz), 280–283.

Congedo, M., Rodrigues, P. L. C., and Jutten, C. (2019). "The Riemannian minimum distance to means field classifier," in *8th International Brain-Computer Interface Conference, BCI 2019* (Graz).

Fahimi, F., Zhang, Z., Goh, W. B., Lee, T. S., Ang, K. K., and Guan, C. (2019). Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI. *J. Neural Eng.* 16:026007. doi: 10.1088/1741-2552/aaf3f6

Fillard, P., Arsigny, V., Ayache, N., and Pennec, X. (2005). A Riemannian framework for the processing of tensor-valued images. *Lect. Notes Comput. Sci.* 3753, 112–123. doi: 10.1007/11577812_10

Frisoli, A., Loconsole, C., Leonardi, D., Banno, F., Barsotti, M., Chisari, C., et al. (2012). A new gaze-BCI-driven control of an upper limb exoskeleton for rehabilitation in real-world tasks. *IEEE Trans. Syst. Man Cybern. Part C* 42, 1169–1179. doi: 10.1109/TSMCC.2012.2226444

He, H., and Wu, D. (2020). Transfer learning for brain-computer interface: a Euclidean space data alignment approach. *IEEE Trans. Biomed. Eng.* 67, 1–22. doi: 10.1109/TBME.2019.2913914

Holm, S. (1979). Board of the foundation of the Scandinavian journal of statistics a simple sequentially rejective multiple test procedure a simple sequentially rejective multiple test procedure. *Scand. J. Stat.* 6, 65–70.

Jayaram, V., Alamgir, M., Altun, Y., Scholkopf, B., and Grosse-Wentrup, M. (2016). Transfer learning in brain-computer interfaces. *IEEE Comput. Intell. Mag.* 11, 20–31. doi: 10.1109/MCI.2015.2501545

Jayaram, V., and Barachant, A. (2018). MOABB: trustworthy algorithm benchmarking for BCIs. *J. Neural Eng.* 15:066011. doi: 10.1088/1741-2552/aadea0

Li, Y., Pan, J., Wang, F., and Yu, Z. (2013). A hybrid BCI system combining p300 and SSVEP and its application to wheelchair control. *IEEE Trans. Biomed. Eng.* 60, 3156–3166. doi: 10.1109/TBME.2013.2270283

Mane, R., Chouhan, T., and Guan, C. (2020). BCI for stroke rehabilitation: motor and beyond. *J. Neural Eng.* 17:041001. doi: 10.1088/1741-2552/aba162

Mayaud, L., Cabanilles, S., Langenhove, A. V., Congedo, M., Barachant, A., Pouplin, S., et al. (2016). Brain-computer interface for the communication of acute patients: a feasibility study and a randomized controlled trial comparing performance with healthy participants and a traditional assistive device. *Brain Comput. Interfaces* 3, 197–215. doi: 10.1080/2326263X.2016.1254403

Minguillon, J., Lopez-Gordo, M. A., and Pelayo, F. (2017). Trends in EEG-BCI for daily-life: requirements for artifact removal. *Biomed. Signal Process. Control* 31, 407–418. doi: 10.1016/j.bspc.2016.09.005

Nielsen, F., and Bhatia, R. (2013). *Matrix Information Geometry*. Heidelberg. doi: 10.1007/978-3-642-30232-9

Perdikis, S., Tonin, L., Saeedi, S., Schneider, C., and del R. Millán, J. (2018). The cyathlon bci race: Successful longitudinal mutual learning

with two tetraplegic users. *PLoS Biol.* 16:5:e2003787. doi: 10.1371/journal.pbio.2003787

Rezeika, A., Benda, M., Stawicki, P., Gembler, F., Saboor, A., and Volosyak, I. (2018). Brain-computer interface spellers: a review. *Brain Sci.* 8:57. doi: 10.3390/brainsci8040057

Rodrigues, P. L., Congedo, M., and Jutten, C. (2021). Dimensionality transcending: a method for merging BCI datasets with different dimensionalities. *IEEE Trans. Biomed. Eng.* 68, 673–684. doi: 10.1109/TBME.2020.3010854

Rodrigues, P. L. C., Jutten, C., and Congedo, M. (2019). Riemannian procrustes analysis: transfer learning for brain-computer interfaces. *IEEE Trans. Biomed. Eng.* 66, 2390–2401. doi: 10.1109/TBME.2018.2889705

Sun, B., Feng, J., and Saenko, K. (2015). Return of frustratingly easy domain adaptation. *Proc. AAAI Conf. Artif. Intell.* 30:1. doi: 10.1609/aaai.v30i1.10306

Tomioka, R., Aihara, K., and Müller, K.-R. (2006). Logistic regression for single trial EEG classification. *Adv. Neural Inform. Process. Syst.* 19, 1377–1384.

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

Xu, J., Markham, A., Meunier, A., Raggam, P., and Grosse-Wentrup, M. (2021). Distance covariance: a nonlinear extension of Riemannian geometry for EEG-based brain-computer interfacing. *IEEE Int. Conf. Syst. Man Cybern.* 2021, 2000–2005. doi: 10.1109/SMC52423.2021.9658876

Yair, O., Ben-Chen, M., and Talmon, R. (2019). Parallel transport on the cone manifold of spd matrices for domain adaptation. *IEEE Trans. Signal Process.* 67, 1797–1811. doi: 10.1109/TSP.2019.2894801

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

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 Trans. Biomed. Eng.* 65, 1107–1116. doi: 10.1109/TBME.2017.2742541

Zaykin, D. V. (2011). Optimally weighted z-test is a powerful method for combining probabilities in meta-analysis. *J. Evol. Biol.* 24, 1836–1841. doi: 10.1111/j.1420-9101.2011.02297.x

Zhang, K., Xu, G., Chen, L., Tian, P., Han, C. C., Zhang, S., et al. (2020). Instance transfer subject-dependent strategy for motor imagery signal classification using deep convolutional neural networks. *Comput. Math. Methods Med.* 2020:1683013. doi: 10.1155/2020/1683013

Zhang, W., and Wu, D. (2020). Manifold embedded knowledge transfer for brain-computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 1117–1127. doi: 10.1109/TNSRE.2020.2985996



OPEN ACCESS

EDITED BY

Davide Valeriani,
Google, United States

REVIEWED BY

Andreas Kalckert,
University of Skövde, Sweden

*CORRESPONDENCE

David M. Eagleman
✉ davideagleman@stanford.edu

SPECIALTY SECTION

This article was submitted to
Brain-Computer Interfaces,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 27 September 2022

ACCEPTED 23 December 2022

PUBLISHED 13 January 2023

CITATION

Eagleman DM and Perrotta MV (2023)
The future of sensory substitution,
addition, and expansion via haptic
devices.
Front. Hum. Neurosci. 16:1055546.
doi: 10.3389/fnhum.2022.1055546

COPYRIGHT

© 2023 Eagleman and Perrotta. This is
an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction
in other forums is permitted, provided
the original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

The future of sensory substitution, addition, and expansion via haptic devices

David M. Eagleman^{1,2*} and Michael V. Perrotta²

¹Department of Psychiatry, Stanford University School of Medicine, Stanford, CA, United States,

²Neosensory, Palo Alto, CA, United States

Haptic devices use the sense of touch to transmit information to the nervous system. As an example, a sound-to-touch device processes auditory information and sends it to the brain via patterns of vibration on the skin for people who have lost hearing. We here summarize the current directions of such research and draw upon examples in industry and academia. Such devices can be used for sensory substitution (replacing a lost sense, such as hearing or vision), sensory expansion (widening an existing sensory experience, such as detecting electromagnetic radiation outside the visible light spectrum), and sensory addition (providing a novel sense, such as magnetoreception). We review the relevant literature, the current status, and possible directions for the future of sensory manipulation using non-invasive haptic devices.

KEYWORDS

senses, sensory substitution devices, hearing loss, haptics, sound

Introduction

The endeavor of getting information to the brain via unusual channels has a long history. We here concentrate on non-invasive devices that use haptics, or the sense of touch. In recent years, as computing technology has advanced, many haptic-based devices have been developed. We categorize these devices into three groups based on their function: sensory substitution, sensory expansion, and sensory addition.

The key to understanding the success of haptics requires remembering that the brain does not directly hear or see the world. Instead, the neural language is built of electrochemical signals in neurons which build some representation of the outside world. The brain's neural networks take in signals from sensory inputs and extract informationally-relevant patterns. It strives to adjust to whatever information it receives and works to extract what it can. As long as the data reflects some relevant feature about the outside world, the brain works to decode it (Eagleman, 2020). In this sense, the brain can be viewed as a general-purpose computing device: it absorbs the available signals and works to determine how to optimally make use of them.

Sensory substitution

Decades ago, researchers realized that the brain's ability to interpret different kinds of incoming information implied that one might be able to get one sensory channel to

carry another's information (Bach-y-Rita et al., 1969). In a surprising demonstration, Bach-y-Rita et al. placed blind volunteers in a reconfigured dental chair in which a grid of four hundred Teflon tips could be extended and retracted by mechanical solenoids. Over the blind participant a camera was mounted on a tripod. The video stream of the camera was converted into a poking of the tips against the volunteer's back. Objects were passed in front of the camera while blind participants in the chair paid careful attention to the feelings in their backs. Over days of training, they became better at identifying the objects by their feel. The blind subjects learned to distinguish horizontal from vertical from diagonal lines, and more advanced users could learn to distinguish simple objects and even faces—simply by the tactile sensations on their back. Bach-y-Rita's findings suggested that information from the skin can be interpreted as readily (if with lower resolution) as information coming from the eyes, and this demonstration opened the floodgates of sensory substitution (Hatwell et al., 2003; Poirier et al., 2007; Bubic et al., 2010; Novich and Eagleman, 2015; Macpherson, 2018).

The technique improved when Bach-y-Rita and his collaborators allowed the blind user to point the camera, using his own volition to control where the “eye” looked (Bach-y-Rita, 1972, 2004). This verified the hypothesis that sensory input is best learned when one can interact with the world. Letting users control the camera closed the loop between muscle output and sensory input (Hurley and Noë, 2003; Noë, 2004). Perception emerges not from a passive input, but instead as a result of actively exploring the environment and matching particular actions to specific changes in sensory inputs. Whether by moving extraocular muscles (as in the case of sighted people) or arm muscles (Bach-y-Rita's participants), the neural architecture of the brain strives to figure out how the output maps to subsequent input (Eagleman, 2020).

The subjective experience for the users was that objects captured by the camera were felt to be located at a distance instead of on the skin of the back (Bach-y-Rita et al., 2003; Nagel et al., 2005). In other words, it was something like vision: instead of stimulating the photoreceptors, the information stimulated touch receptors on the skin, resulting in a functionally similar experience.

Although Bach-y-Rita's vision-to-touch system was the first to seize the public imagination, it was not the first attempt at sensory substitution. In the early 1960s, Polish researchers had passed visual information via touch, building a system of vibratory motors mounted on a helmet that “drew” the images on the head through vibrations [the *Elektroftalm*; (Starkiewicz and Kuliszewski, 1963)]. Blind participants were able to navigate specially prepared rooms that were painted to enhance the contrast of door frames and furniture edges. Unfortunately, the device was heavy and would get hot during use, and thus was not market-ready—but the proof of principle was there.

These unexpected approaches worked because inputs to the brain (such as photons at the eyes, air compression waves at the ears, pressure on the skin) are all converted into electrical signals. As long as the incoming spikes carry information that represents something important about the outside world, the brain will attempt to interpret it.

In the 1990s, Bach-y-Rita et al. sought ways to go smaller than the dental chair. They developed a small device called the BrainPort (Bach-y-Rita et al., 2005; Nau et al., 2015; Stronks et al., 2016). A camera is attached to the forehead of a blind person, and a small grid of electrodes is placed on the tongue. The “Tongue Display Unit” of the BrainPort uses a grid of stimulators over three square centimeters. The electrodes deliver small shocks that correlate with the position of pixels, feeling something like Pop Rocks candy in the mouth. Bright pixels are encoded by strong stimulation at the corresponding points on the tongue, gray by medium stimulation, and darkness by no stimulation. The BrainPort gives the capacity to distinguish visual items with a visual acuity that equates to about 20/800 vision (Sampaio et al., 2001). While users report that they first perceive the tongue stimulation as unidentifiable edges and shapes, they eventually learn to recognize the stimulation at a deeper level, allowing them to discern qualities such as distance, shape, direction of movement, and size (Stronks et al., 2016).

The tongue provides an excellent brain-machine interface because it is densely packed with touch receptors (Bach-y-Rita et al., 1969; Bach-y-Rita, 2004). When brain imaging is performed on trained subjects (blind or sighted), the motion of electrotactile shocks across the tongue activates the MT+ area of the visual cortex, an area which is normally involved in visual motion (Merabet et al., 2009; Amedi et al., 2010; Matteau et al., 2010).

Of particular interest is the subjective experience. The blind participant Roger Behm describes the experience of the BrainPort:

Last year, when I was up here for the first time, we were doing stuff on the table, in the kitchen. And I got kind of... a little emotional, because it's 33 years since I've seen before. And I could reach out and see the different-sized balls. I mean I visually see them. I could reach out and grab them—not grope or feel for them—pick them up, and see the cup, and raise my hand and drop it right in the cup (Bains, 2007).

Tactile input can work on many locations on the body. For example, the Forehead Retina System converts a video stream into a small grid of touch on the forehead (Kajimoto et al., 2006). Another device hosts a grid of vibrotactile actuators on the abdomen, which use intensity to represent distance to the nearest surfaces. Researchers used this device to demonstrate that blind participants' walking trajectories are not preplanned, but instead emerge dynamically as the tactile information streams in Lobo et al. (2017, 2018).

As 5% of the world has disabling hearing loss, researchers have recently sought to build sensory substitution for the deaf (Novich and Eagleman, 2015). Assimilating advances in high-performance computing into a sound-to-touch sensory-substitution device worn under the shirt, Novich and Eagleman (2015) built a vest that captured sound around the user and mapped it onto vibratory motors on the skin, allowing users to feel the sonic world around them. The theory was to transfer the function of the inner ear (breaking sounds into different frequencies and sending the data to the brain) to the skin.

Does the skin have enough bandwidth to transmit all the information of sound? After all, the cochlea is an exquisitely specialized structure for capturing sound frequencies with high fidelity, while the skin is focused on other measures and has poor spatial resolution. Conveying a cochlear-level of information through the skin would require several thousand vibrotactile motors—too many to fit on a person. However, by compressing the speech information, just a few motors suffice (Koffler et al., 2015; Novich and Eagleman, 2015). Such technology can be designed in many different form factors, such as a chest strap for children and a wristband with vibratory motors (Figure 1).

On the first day of wearing the wristband, users are at the very least able to use the vibrations as cues that a noise is happening. Users quickly learn to use the vibrations to differentiate sounds, such as a dog barking, a faucet running, a doorbell ringing, or someone calling their name. A few days into wearing the wristband, users report that the conscious perception of the vibrations fades into the background but still aids them in knowing what sounds are nearby. Users' ability to differentiate patterns of vibrations improves over time. Figure 1C shows performance scores over time on a three-alternative forced choice paradigm, where the three choices were picked at random from a list of 14 environmental sounds. One environmental sound was presented as vibrations on the wristband and the user had to choose which sound they felt (Perrotta et al., 2021).

Moreover, after several months users develop what appears to be a direct subjective experience of external sound. After 6 months, one user reported that he no longer has a sensation of buzzing followed by an interpretation of those vibrations, but instead, "I perceive the sound in my head" (personal interview). This was a subjective report and qualia are not possible to verify; nonetheless we found the claim sufficiently interesting to note here.

The idea of converting touch into sound is not new (Traunmüller, 1980; Cholewiak and Sherrick, 1986; Weisenberger et al., 1991; Summers and Gratton, 1995; Galvin et al., 2001; Reed and Delhorne, 2005). In 1923, Robert Gault, a psychologist at Northwestern University, heard about a deaf and blind ten-year-old girl who claimed to be able to feel sound through her fingertips. Skeptical, he ran experiments. He stopped up her ears and wrapped her head in a woolen blanket (and verified on his graduate student that this prevented the

ability to hear). She put her finger against the diaphragm of a "portophone" (a long hollow tube), and Gault sat in a closet and spoke through it. Her only ability to understand what he was saying was from vibrations on her fingertip. He reports,

After each sentence or question was completed her blanket was raised and she repeated to the assistant what had been said with but a few unimportant variations.... I believe we have here a satisfactory demonstration that she interprets the human voice through vibrations against her fingers.

Gault mentions that his colleague has succeeded at communicating words through a thirteen-foot-long glass tube. A trained participant, with stopped-up ears, could put his palm against the end of the tube and identify words that were spoken into the other end. With these sorts of observations, researchers have attempted to make sound-to-touch devices, but until recent decades the machinery was too large and computationally weak to make for a practical device.

Similarly, the Tadoma method, developed in the 1930s, allows people who are deaf and blind to understand the speech of another person by placing a hand over the face and neck of the speaker. The thumb rests lightly on the lips and the fingers fan out to cover the neck and cheek, allowing detection of moving lips, vibrating vocal cords, and air coming out of the nostrils. Thousands of deaf and blind children have been taught this method and have obtained proficiency at understanding language almost to the point of those with hearing, all through touch (Alcorn, 1945).

In the 1970s, deaf inventor Dimitri Kanevsky developed a two-channel vibrotactile device, one of which captures the envelope of low frequencies, and the other high. Two vibratory motors sit on the wrists. By the 1980s, similar inventions in Sweden and the United States were proliferating. The problem was that all these devices were too large, with too few motors (typically just one) to make an impact. Due to computational limitations in previous years, earlier attempts at sound-to-touch substitution relied on band-pass filtering audio and playing this output to the skin over vibrating solenoids. The solenoids operated at a fixed frequency of less than half the bandwidth of some of these band-passed channels, leading to aliasing noise. Further, multichannel versions of these devices were limited in the number of actuators due to battery size and capacity constraints. With modern computation, the desired mathematical transforms can be performed in real time, at little expense, and without the need of custom integrated circuits, and the whole device can be made as an inexpensive, wearable computing platform.

A wrist-worn sound-to-touch sensory substitution device was recently shown in brain imaging to induce activity in both somatosensory and auditory regions, demonstrating that the brain rapidly recruits existing auditory processing areas to aid in the understanding of the touch (Malone et al., 2021).

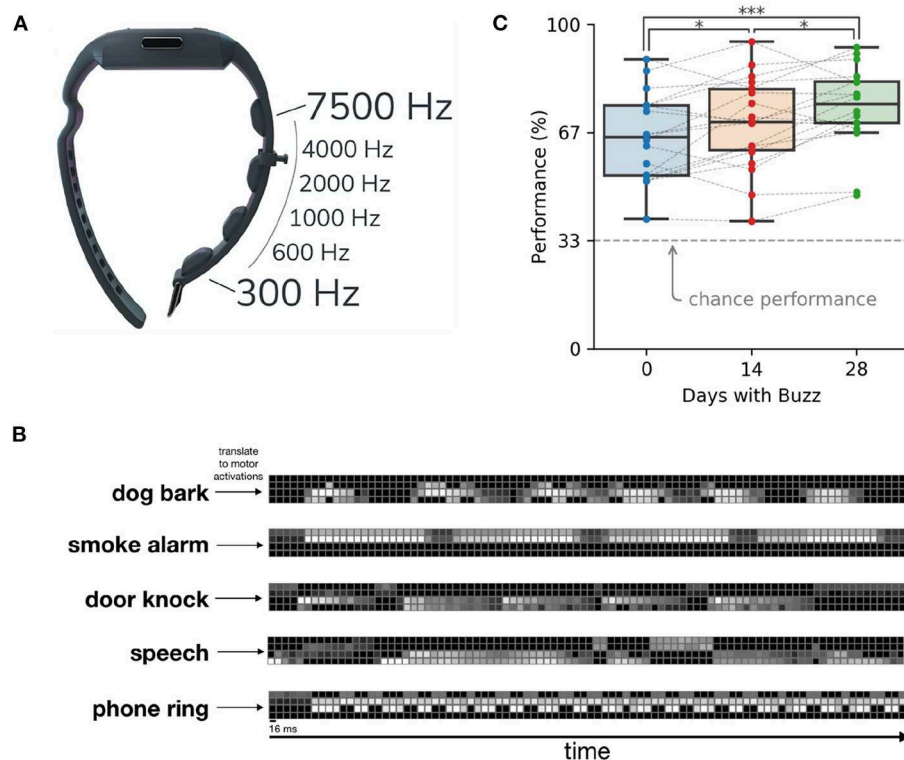


FIGURE 1

A sensory substitution wristband for deafness (the Neosensory Buzz). (A) Four vibratory motors in a wristband transfer information about sound levels in different frequency bands. (B) Sample sounds translated into patterns of vibration on the motors of the wristband. Brighter colors represent higher intensity of the motor. (C) Participant performance on sound identification improves through time. Data reprinted from Perrotta et al. (2021).

There are cost advantages to a sensory substitution approach. Cochlear implants typically cost around \$100,000 for implantation (Turchetti et al., 2011). In contrast, haptic technologies can address hearing loss for some hundreds of dollars. Implants also require an invasive surgery, while a vibrating wristband is merely strapped on like a watch.

There are many reasons to take advantage of the system of touch. For example, people with prosthetic legs have difficulty learning how to walk with their new prosthetics because of a lack of proprioception. To allow participants to understand the position of an artificial limb, other sensory devices can channel the information. For example, research has shown improvements in stair descent for lower limb amputees using haptic sensory substitution devices (Sie et al., 2017), and haptic sensory substitution devices have also been created for providing sensory feedback for upper limb prosthetics (Cipriani et al., 2012; Antfolk et al., 2013; Rombokas et al., 2013). Some sensory substitution devices for upper limb amputees use electrotactile stimulation instead of vibrotactile stimulation, targeting different receptors in the skin (Saleh et al., 2018).

This same technique can be used for a person with a real leg that has lost sensation—as happens in Parkinson's disease

or peripheral neuropathy. In unpublished internal experiments, we have successfully piloted a solution that used sensors in a sock to measure motion and pressure and fed the data into the vibrating wristband. By this technique, a person understands where her foot is, whether her weight is on it, and whether the surface she's standing on is even. A recent systematic review synthesizing the findings of nine randomized controlled trials showed that sensory substitution devices are effective in improving balance measures of neurological patient populations (Lynch and Monaghan, 2021).

Touch can also be used to address problems with balance. This has been done with the BrainPort tongue display (Tyler et al., 2003; Danilov et al., 2007): the head orientation was fed to the BrainPort tongue grid: when the head was straight up, the electrical stimulation was felt in the middle of the tongue grid; when the head tilted forward, the electrical signal moved toward the tip of the tongue; when the head tilted back, the stimulation moved toward the rear; side-to-side tilts were encoded by left and right movement of the electrical signal. In this way, a person who had lost all sense of which way her head was oriented could feel the answer on her tongue. Of note, the residual benefits extended even after taking off the device. Users' brains figured

out how to take residual vestibular signals as well as existing visual and proprioceptive signals and strengthen them with the guidance of the helmet. After several months of using the helmet, many participants were able to reduce the frequency of their usage.

We have developed a similar experimental system that is currently unpublished but bears mentioning for illustration purposes. In a balance study underway at Stanford University, we use a vibratory wristband in combination with a 9-axis inertial measurement unit (IMU) that is clipped to the collar of a user. The IMU outputs an absolute rotation relative to a given origin, which is set as the device's position when the user is standing upright. The pitch and roll of the rotation are mapped to vibrations on the wristband to provide additional balance information to the user's brain: the more one tilts away from upright, the higher the amplitude of vibrations one feels on the wrist. The direction of the tilt (positive or negative pitch or roll) is mapped to the vibration location on the wrist. Results of this approach will be published in the future.

Besides the basic five senses, more complex senses can be aided with sensory substitution devices. People with autism spectrum disorder often have a decreased ability to detect emotion in others; this, in one preliminary project, machine learning algorithms classify the emotional states detected in speech and communicate these emotional states to the brain via vibrations on the wrist. Currently, the machine learning algorithm detects and communicates how much someone's speech matches seven different emotions (neutral, surprise, disgust, happiness, sadness, fear, and anger) and communicates that to the wearer (ValenceVibrations.com).

Sensory substitution opens new opportunities to compensate for sensory loss. However, similar devices can move past compensation and instead build on top of normal senses—we call these devices sensory expansion devices. Instead of filling in gaps for someone with a sensory deficit, these expand the unhindered senses to be better, wider, or faster.

Sensory expansion

Many examples of sensory expansion have been demonstrated in animals. For example, mice and monkeys can be moved from color-blindness to color vision by genetically engineering photoreceptors to contain human photopigment ([Jacobs et al., 2007](#); [Mancuso et al., 2009](#)). The research team injected a virus containing the red-detecting opsin gene behind the retina. After 20 weeks of practice, the monkeys could use the color vision to discriminate previously indistinguishable colors.

In a less invasive example, we created a sensory expansion device by connecting a vibrating wristband to a near-wavelength infrared sensor and an ultraviolet sensor. Although our eyes capture only visible light, the frequencies of the electromagnetic spectrum adjacent to visible light are in fact visible to a variety

of animals. For instance, honeybees can see ultraviolet patterns on flowers ([Silberglied, 1979](#)). By capturing the intensity of light in these ranges and mapping those intensities to vibrations, a user can pick up on information in these invisible light regions without gene editing or retinal implants. In this way, a wrist-worn device can expand vision beyond its natural capabilities. One of us (DME) wore an infrared bolometer connected to a haptic wristband and was able to easily detect infrared cameras in the darkness ([Eagleman, 2020](#)).

To illustrate the breadth of possibilities, it bears mention that we have performed an unpublished preliminary experiment with blindness. Using lidar (light detection and ranging), we tracked the position of every moving object in an office space—in this case, humans moving around. We connected the data from the lidar sensors to our vibrating vest, such that the vest vibrated to tell the wearer if they were approaching an obstacle like a wall or chair, where there were people nearby, and what direction they should move to most quickly reach a target destination. We tested this sensory expansion device with a blind participant. He wore the vest and could feel the location of objects and people around him as well as the quickest path to a desired destination (such as a conference room). Interestingly, there was no learning curve: he immediately understood how to use the vibrations to navigate without colliding into objects or people. Although sensory substitution devices can fill the gap left by vision impairment, this device did more than that—it offered an expanded, 360° sense of space. A sighted person could also wear this device to expand their sense of space, allowing them to know what objects or people are behind them. Because this device does more than alleviate a sensory loss, it is an example of sensory expansion.

Haptic sensory expansion is not limited to vision. Devices from hearing aids to the Buzz can reach beyond the normal hearing scale—for example, into the ultrasonic range (as heard by cats or bats), or the infrasonic (as heard by elephants) ([Wolbring, 2013](#)).

The sense of smell can also be benefited by sensory expansion. To illustrate an unpublished possibility, imagine converting the data from an array of molecular detectors into haptic signals. While this is unproven, the goal should be clear: for a person to access a new depth of odor detection, beyond the natural sensory acuity of human smell.

One can also detect temperature via sensory expansion. In preliminary experiments, participants use an array of mid-wavelength infrared sensors to detect the temperature of nearby objects and translate the data to vibrations on the wrist. The wearer learns to interpret the vibrations as a sense of temperature, but one that does not stop at the skin—instead, their sense of temperature has expanded to include objects in the surrounding environment.

For the purpose of illustrating the width of possibilities, we note that internal signals in the body—such as the sense of one's own blood sugar levels—can be easily expanded by combining

easily-obtainable technologies. For example, continuous glucose monitoring devices allow users to look at their level of blood sugar at any point; however, the user still must pull out a cell phone to consult an app. By connecting a haptic device to a continuous glucose monitoring device, one could create a sensory expansion device that allows users to have continuous access to their blood sugar levels without having to visually attend to a screen.

More broadly, one could also make a device to expand one's sense of a partner's wellbeing. By connecting sensors that detect a partner's breathing rate, temperature, galvanic skin response, heart rate, and more, a haptic device could expand the user's information flow to allow them to feel their partner's internal signals. Interestingly, this sense need not be limited by proximity: the user can sense how their partner is feeling even from across the country, so long as they have an internet connection.

An important question for any haptic sensory device is whether the user is gaining a new sensory experience or is instead consciously processing the incoming haptic information. In the former case, the subjective experience of a temperature sensing wristband would be similar to the subjective experience of touching a hot stove (without needing to touch or be near the surface), while the latter case would be closer to feeling an alert on the wrist that warns of a hot surface. As mentioned above, previous work has shown evidence for a direct subject experience of a new sense, whether from brain imaging or through subjective questionnaires (Bach-y-Rita et al., 2003; Nagel et al., 2005). Until this investigation is done on each new sensory device, it is unknown whether the device is providing a new sense or rather a signal that can be consciously perceived as touch.

While these devices all expand on one existing sense or another, other devices can go further, representing entirely new senses. These devices form the third group: sensory addition.

Sensory addition

Due to the brain's remarkable flexibility, there is the possibility of leveraging entirely new data streams directly into perception (Hawkins and Blakeslee, 2007; Eagleman, 2015).

One increasingly common example is the implantation of small neodymium magnets into the fingertips. By this method, "biohackers" can haptically feel magnetic fields. The magnets tug when exposed to electromagnetic fields, and the nearby touch nerves register this. Information normally invisible to humans is now streamed to the brain via the sensory nerves from the fingers. People report that detecting electromagnetic fields (e.g., from a power transformer) is like touching an invisible bubble, one with a shape that can be assessed by moving one's hand around (Dvorsky, 2012). A world is detectable that

previously was not: palpable shapes live around microwave ovens, computer fans, speakers, and subway power transformers.

Can haptic devices achieve the same outcome without implanting magnets into the fingertips? One developer created a sensory addition device using a haptic wristband that translates electromagnetic fields into vibrations (details of the project at neosensory.com/developers). Not only is such an approach less invasive, it is also more customizable. Instead of just feeling the presence of an electromagnetic field, this device decomposes the frequency of an alternating current signal and presents the intensity of different parts of the spectrum via different vibrating motors. Thus, an electrician can add this new sense to their perception, knowing the frequency and intensity of electric signals flowing through live wires.

What if you could detect not only the magnetic field around objects but also the one around the planet—as many animal species do? Researchers at Osnabrück University devised a belt called the feelSpace to allow humans to tap into that signal. The belt is ringed with vibratory motors, and the motor pointed to the north buzzes. As you turn your body, you always feel the buzzing in the direction of magnetic north.

At first, it feels like buzzing, but over time it becomes spatial information: a feeling that north is *there* (Kaspar et al., 2014). Over several weeks, the belt changes how people navigate: their orientation improves, they develop new strategies, they gain a higher awareness of the relationship between different places. The environment seems more ordered. Relationships between places can be easily remembered.

As one subject described the experience, "The orientation in the cities was interesting. After coming back, I could retrieve the relative orientation of all places, rooms and buildings, even if I did not pay attention while I was actually there" (Nagel et al., 2005). Instead of thinking about a sequence of cues, they thought about the situation globally. Another user describes how it felt: "It was different from mere tactile stimulation, because the belt mediated a spatial feeling.... I was intuitively aware of the direction of my home or of my office." In other words, his experience is not of sensory substitution, nor is it sensory expansion (making your sight or hearing better). Instead, it's a sensory addition. It's a new kind of human experience. The user goes on:

During the first 2 weeks, I had to concentrate on it; afterwards, it was intuitive. I could even imagine the arrangement of places and rooms where I sometimes stay. Interestingly, when I take off the belt at night I still feel the vibration: When I turn to the other side, the vibration is moving too—this is a fascinating feeling! (Nagel et al., 2005).

After users take off the feelSense belt, they often report that they continue having a better sense of orientation for some time. In other words, the effect persists even without wearing the device. As with the balance helmet, weak internal signals can get

strengthened when an external device confirms them. (Note that one won't have to wear a belt for long; researchers have recently developed a thin electronic skin—essentially a little sticker on the hand—that indicates north; see [Bermúdez et al., 2018](#)).

Other projects tackle tasks that require a great deal of cognitive load. For example, a modern aircraft cockpit is packed with visual instruments. With a sensory addition device, a pilot can feel the high-dimensional stream of data instead of having to read all the data visually. The North Atlantic Treaty Organization (NATO) released a report on haptic devices developed to help pilots navigate in low-visibility settings ([Van Erp and Self, 2008](#)). Similarly, [Fellah and Guiatni \(2019\)](#) developed a haptic sensory substitution device to give pilots access to the turn rate angle, climb angle, and flight control warning messages via vibrations.

As a related example, researchers at our laboratory are piloting a system to allow doctors to sense the vitals of a patient without having to visually consult a variety of monitors. This device connects a haptic wristband to an array of sensors that measure body temperature, blood oxygen saturation, heart rate, and heart rate variability. Future work will optimize how these data streams are presented to the doctor and with what resolution. Both psychophysical testing and understanding user needs will shape this optimized mapping (for example, how much resolution can a doctor learn to feel in the haptic signal, and what is the smallest change in blood pressure that should be discernible via the device?).

Finally, it is worth asking whether haptic devices are optimal for sending data streams to the brain. After all, one could leverage a higher-resolution sense, such as vision or audition, or perhaps use multiple sensory modalities. This is an open question for the future; however, haptics is advantageous due simply to the fact that vision and hearing are necessary for so many daily tasks. Skin is a high-bandwidth, mostly unused information channel—and therefore its almost-total availability makes it an attractive target for new data streams.

Conclusion

We have reviewed some of the projects and possibilities of non-invasive, haptic devices for passing new data streams to the brain. The chronic rewiring of the brain gives it tremendous flexibility: it dynamically reconfigures itself to absorb and interact with data. As a result, electrical grids can come to feed

visual information via the tongue, vibratory motors can feed hearing via the skin, and cell phones can feed video streams via the ears. Beyond sensory substitution, such devices can be used to endow the brain with new capacities, as we see with sensory expansion (extending the limits of an already-existing sense) and sensory addition (using new data streams to create new senses). Haptic devices have moved rapidly from computer-laden cabled devices to wireless wearables, and this progress, more than any change in the fundamental science, will increase their usage and study.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

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

Author contributions

DE wrote the manuscript. MP engineered many of the devices described. Both authors contributed to the article and approved the submitted version.

Conflict of interest

DE and MP were employed by Neosensory.

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.

References

- Alcorn, S. (1945). Development of the Tadoma method for the deaf-blind. *J. Except. Children*. 77, 247–57. doi: 10.1177/001440294501100407
- Amedi, A., Raz, N., Azulay, H., Malach, R., and Zohary, E. (2010). Cortical activity during tactile exploration of objects in blind and sighted humans. *Restorat. Neurol. Neurosci.* 28, 143–156. doi: 10.3233/RNN-2010-0503

- Antfolk, C., D'Alonzo, M., Rosén, B., Lundborg, G., Sebelius, F., and Cipriani, C. (2013). Sensory feedback in upper limb prosthetics. *Expert Rev. Med. Dev.* 10, 45–54. doi: 10.1586/erd.12.68
- Bach-y-Rita, P. (1972). *Brain Mechanisms in Sensory Substitution*. Berlin: Springer.
- Bach-y-Rita, P. (2004). Tactile sensory substitution studies. *Ann New York Acad Sci.* 1013, 83–91. doi: 10.1196/annals.1305.006
- Bach-y-Rita, P., Collins, C. C., Saunders, F. A., White, B., and Scadden, L. (1969). Vision substitution by tactile image projection. *Nature*. 221, 963–964. doi: 10.1038/221963a0
- Bach-y-Rita, P., Danilov, Y., Tyler, M. E., and Grimm, R. J. (2005). *Late Human Brain Plasticity: Vestibular Substitution with a Tongue BrainPort Human-Machine Interface*. Intellectica. Revue de l'Association Pour La Recherche Cognitive. doi: 10.3406/intel.2005.1362
- Bach-y-Rita, P., Tyler, M. E., and Kaczmarek, K. A. (2003). Seeing with the Brain. *Int J Hum Comput Interact.* 15, 285–295. doi: 10.1207/S15327590IJHC1502_6
- Bains, S. (2007). *Mixed Feelings*. Wired. Conde Nast. Available online at: <https://www.wired.com/2007/04/esp/>
- Bermúdez, G. S. C., Fuchs, H., Bischoff, L., Fassbender, J., and Makarov, D. (2018). Electronic-skin compasses for geomagnetic field-driven artificial magnetoreception and interactive electronics. *Nat. Electron.* 1, 589–95. doi: 10.1038/s41928-018-0161-6
- Bubic, A., Striem-Amit, E., and Amedi, A. (2010). *Large-Scale Brain Plasticity Following Blindness and the Use of Sensory Substitution Devices. Multisensory Object Perception in the Primate Brain*. Berlin: Springer. doi: 10.1007/978-1-4419-5615-6_18
- Cholewiak, R. W., and Sherrick, C. E. (1986). Tracking skill of a deaf person with long-term tactile aid experience: a case study. *J. Rehabil. Res. Develop.* 23, 20–26.
- Cipriani, C., D'Alonzo, M., and Carrozza, M. C. (2012). A miniature vibrotactile sensory substitution device for multifingered hand prosthetics. *IEEE Trans. Bio-Med. Eng.* 59, 400–408. doi: 10.1109/TBME.2011.2173342
- Danilov, Y. P., Tyler, M. E., Skinner, K. L., Hogle, R. A., and Bach-y-Rita, P. (2007). Efficacy of electrotactile vestibular substitution in patients with peripheral and central vestibular loss. *J. Vestibular Res. Equilib. Orient.* 17, 119–130. doi: 10.3233/VES-2007-172-307
- Dvorsky, G. (2012). *What Does the Future Have in Store for Radical Body Modification?*. Gizmodo. Available online at: <https://gizmodo.com/what-does-the-future-have-in-store-for-radical-body-mod-5944883> (accessed September 20, 2012).
- Eagleman, D. (2015). “Can We Create New Senses for Humans?” TED. Available online at: https://www.ted.com/talks/david_eagleman_can_we_create_new_senses_for_humans?language=en
- Eagleman, D. (2020). *Livewired: The Inside Story of the Ever-Changing Brain*. Vintage.
- Fellah, K., and Guiatni, M. (2019). Tactile display design for flight envelope protection and situational awareness. *IEEE Trans. Haptics* 12, 87–98. doi: 10.1109/TOH.2018.2865302
- Galvin, K. L., Ginis, J., Cowan, R. S. C., Blamey, P. J., and Clark, G. M. (2001). A comparison of a new prototype tickle talkerm with the tactaid 7. *Aust. New Zealand J. Audiol.* 23, 18–36. doi: 10.1375/audi.23.1.18.31095
- Hatwell, Y., Streri, A., and Gentaz, E. (2003). *Touching for Knowing: Cognitive Psychology of Haptic Manual Perception*. Amsterdam: John Benjamins Publishing. doi: 10.1075/aicr.53
- Hawkins, J., and Blakeslee, S. (2007). *On Intelligence: How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines*. New York, NY: Macmillan.
- Hurley, S., and Noë, A. (2003). Neural plasticity and consciousness: reply to block. *Trends Cognit. Sci.* 7, 342. doi: 10.1016/S1364-6613(03)00165-7
- Jacobs, G. H., Williams, G. A., Cahill, H., and Nathans, J. (2007). Emergence of novel color vision in mice engineered to express a human cone photopigment. *Science* 315, 1723–1725. doi: 10.1126/science.1138838
- Kajimoto, H., Kanno, Y., and Tachi, S. (2006). “Forehead electro-tactile display for vision substitution,” in *Proc. EuroHaptics*, 11. Citeseer.
- Kaspar, K., König, S., Schwandt, J., and König, P. (2014). The experience of new sensorimotor contingencies by sensory augmentation. *Consciousness Cognit.* 28, 47–63. doi: 10.1016/j.concog.2014.06.006
- Koffler, T., Ushakov, K., and Avraham, K. B. (2015). Genetics of hearing loss: syndromic. *Otolaryngol Clin North Am.* 48, 1041–1061. doi: 10.1016/j.otc.2015.07.007
- Lobo, L., Higuera-Herbada, A., Travieso, D., Jacobs, D. M., Rodger, M., and Craig, C. M. (2017). “Sensory substitution and walking toward targets: an experiment with blind participants,” in *Studies in Perception and Action XIV: Nineteenth International Conference on Perception and Action* (Milton Park: Psychology Press). p. 35.
- Lobo, L., Travieso, D., Jacobs, D. M., Rodger, M., and Craig, C. M. (2018). Sensory substitution: using a vibrotactile device to orient and walk to targets. *J. Exp. Psychol. Appl.* 24, 108–124. doi: 10.1037/xap0000154
- Lynch, P., and Monaghan, K. (2021). Effects of sensory substituted functional training on balance, gait, and functional performance in neurological patient populations: a systematic review and meta-analysis. *Heliyon* 7, e08007. doi: 10.1016/j.heliyon.2021.e08007
- Macpherson, F. (2018). *Sensory Substitution and Augmentation: An Introduction*. Sensory Substitution and Augmentation. doi: 10.5871/bacad/9780197266441.003.0001
- Malone, P. S., Eberhardt, S. P., Auer, E. T., Klein, R., Bernstein, L. E., and Riesenhuber, M. (2021). *Neural Basis of Learning to Perceive Speech through Touch Using an Acoustic-to-Vibrotactile Speech Sensory Substitution*. bioRxiv. doi: 10.1101/2021.10.24.465610
- Mancuso, K., Hauswirth, W. W., Li, Q., Connor, T. B., Kuchenbecker, J. A., Mauck, M. C., et al. (2009). Gene therapy for red-green colour blindness in adult primates. *Nature* 461, 784–787. doi: 10.1038/nature08401
- Matteau, I., Kupers, R., Ricciardi, E., Pietrini, P., and Ptito, M. (2010). Beyond visual, aural and haptic movement perception: hMT+ Is activated by electrotactile motion stimulation of the tongue in sighted and in congenitally blind individuals. *Brain Res. Bull.* 82, 264–270. doi: 10.1016/j.brainresbull.2010.05.001
- Merabet, L. B., Battelli, L., Obretenova, S., Maguire, S., Meijer, P., and Pascual-Leone, A. (2009). Functional recruitment of visual cortex for sound encoded object identification in the blind. *Neuroreport* 20, 132–138. doi: 10.1097/WNR.0b013e328323104dc
- Nagel, S. K., Carl, C., Kringe, T., Martin, R., and König, P. (2005). Beyond sensory substitution—learning the sixth sense. *J. Neural Eng.* 2, R13–26. doi: 10.1088/1741-2560/2/4/R02
- Nau, A. C., Pintar, C., Arnoldussen, A., and Fisher, C. (2015). Acquisition of visual perception in blind adults using the brainport artificial vision device. *Am. J. Occup. Therapy* 69, 1–8. doi: 10.5014/ajot.2015.011809
- Noë, A. (2004). *Action in Perception*. Cambridge: MIT Press.
- Novich, S. D., and Eagleman, D. M. (2015). Using space and time to encode vibrotactile information: toward an estimate of the skin's achievable throughput. *Exp. Brain Res.* 23, 2777–88. doi: 10.1007/s00221-015-4346-1
- Perrotta, M. V., Asgeirsdottir, T., and Eagleman, D. M. (2021). Deciphering sounds through patterns of vibration on the skin. *Neuroscience* 458, 77–86. doi: 10.1016/j.neuroscience.2021.01.008
- Poirier, C., Volder, A. G. D., and Scheiber, C. (2007). What neuroimaging tells us about sensory substitution. *Neurosci. Biobehav. Rev.* 31, 1064–70. doi: 10.1016/j.neubiorev.2007.05.010
- Reed, C. M., and Delhorne, L. A. (2005). Reception of environmental sounds through cochlear implants. *Ear Hear.* 26, 48–61. doi: 10.1097/00003446-200502000-00005
- Rombokas, E., Stepp, C. E., Chang, C., Malhotra, M., and Matsuoka, Y. (2013). Vibrotactile sensory substitution for electromyographic control of object manipulation. *IEEE Trans. Bio-Med. Eng.* 60, 2226–2232. doi: 10.1109/TBME.2013.2252174
- Saleh, M., Ibrahim, A., Ansovini, F., Mohanna, Y., and Valle, M. (2018). “Wearable system for sensory substitution for prosthetics,” in *2018 New Generation of CAS (NGCAS)*. p. 110–13. doi: 10.1109/NGCAS.2018.8572173
- Sampaio, E., Maris, S., and Bach-y-Rita, P. (2001). Brain plasticity: ‘visual’ acuity of blind persons via the tongue. *Brain Res.* 908, 204–207. doi: 10.1016/S0006-8993(01)02667-1
- Sie, A., Realmuto, J., and Rombokas, E. (2017). “A lower limb prosthesis haptic feedback system for stair descent,” in *2017 Design of Medical Devices Conference*. American Society of Mechanical Engineers Digital Collection. doi: 10.1115/DMD2017-3409
- Silberglied, R. E. (1979). Communication in the ultraviolet. *Ann. Rev. Ecol. Syst.* 10, 373–398. doi: 10.1146/annurev.es.10.110179.002105
- Starkiewicz, W., and Kuliszewski, T. (1963). The 80-channel elektroftalm. *Proc. Int. Congress Technol. Blindness* 1, 157.
- Stronks, H. C., Mitchell, E. B., Nau, A. C., and Barnes, N. (2016). Visual task performance in the blind with the brainport v100 vision aid. *Expert Rev. Med. Dev.* 13, 919–931. doi: 10.1080/17434440.2016.1237287

Summers, I. R., and Gratton, D. A. (1995). Choice of speech features for tactile presentation to the profoundly deaf. *IEEE Trans. Rehabil. Eng.* 3, 117–121. doi: 10.1109/86.372901

Traunmüller, H. (1980). The sentiphone: a tactual speech communication aid. *J. Commun. Disorders* 13, 183–193. doi: 10.1016/0021-9924(80)90035-0

Turchetti, G., Bellelli, S., Palla, I., and Forli, F. (2011). Systematic review of the scientific literature on the economic evaluation of cochlear implants in paediatric patients. *Acta Otorhinolaryngol. Ital.* 31, 311–318.

Tyler, M., Danilov, Y., and Bach-Y-Rita, P. (2003). Closing an open-loop control system: vestibular substitution through the tongue. *J. Integrat. Neurosci.* 2, 159–164. doi: 10.1142/S0219635203000263

Van Erp, J. B., and Self, B. P. (2008). *Tactile Displays for Orientation, Navigation and Communication in Air, Sea and Land Environments*. North Atlantic Treaty Organisation, Research and Technology Organisation.

Weisenberger, J. M., Craig, J. C., and Abbott, G. D. (1991). Evaluation of a principal-components tactile aid for the hearing-impaired. *J. Acoust. Soc. Am.* 90, 1944–1957. doi: 10.1121/1.401674

Wolbring, G. (2013). Hearing beyond the normal enabled by therapeutic devices: the role of the recipient and the hearing profession. *Neuroethics* 6, 607–616. doi: 10.1007/s12152-011-9120-x

Frontiers in Human Neuroscience

Bridges neuroscience and psychology to
understand the human brain

The second most-cited journal in the field of
psychology, that bridges research in psychology
and neuroscience to advance our understanding
of the human brain in both healthy and diseased
states.

Discover the latest Research Topics

[See more →](#)

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact



Frontiers in Human Neuroscience

