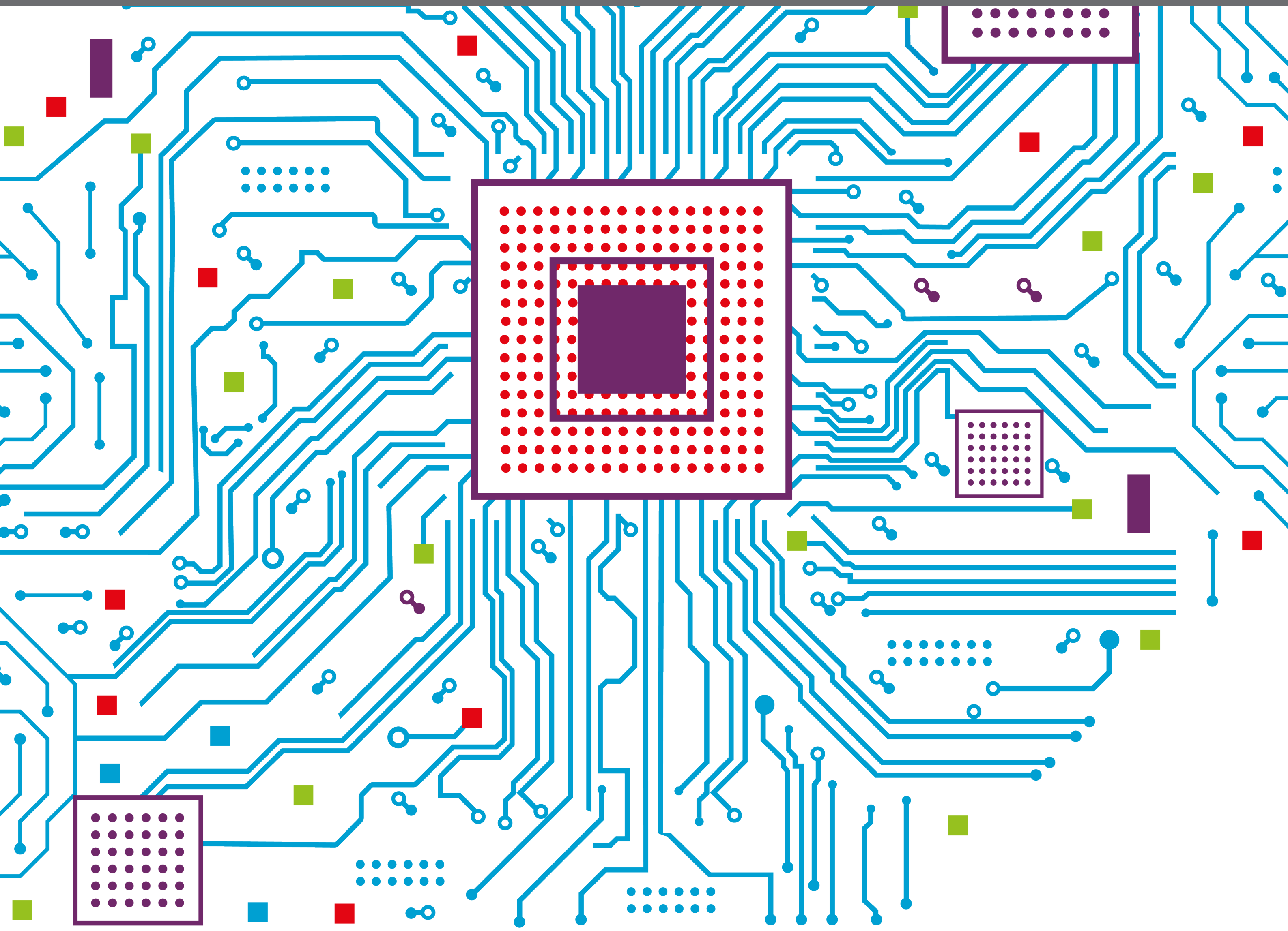
BRAIN-COMPUTER INTERFACES FOR NON-CLINICAL (HOME, SPORTS, ART, ENTERTAINMENT, EDUCATION, WELL-BEING) APPLICATIONS

EDITED BY: Anton Nijholt, Jose Luis Contreras-Vidal, Camille Jeunet and Aleksander Väljamäe

PUBLISHED IN: Frontiers in Computer Science and Frontiers in Human Neuroscience



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88974-717-7
DOI 10.3389/978-2-88974-717-7

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

BRAIN-COMPUTER INTERFACES FOR NON-CLINICAL (HOME, SPORTS, ART, ENTERTAINMENT, EDUCATION, WELL-BEING) APPLICATIONS

Topic Editors:
Anton Nijholt, University of Twente, Netherlands
Jose Luis Contreras-Vidal, University of Houston, United States
Camille Jeunet, UMR5287 Institut de Neurosciences Cognitives et Intégratives d'Aquitaine (INCIA), France
Aleksander Väljamäe, Tallinn University, Estonia

Citation: Nijholt, A., Contreras-Vidal, J. L., Jeunet, C., Väljamäe, A., eds. (2022). Brain-Computer Interfaces for Non-clinical (Home, Sports, Art, Entertainment, Education, Well-being) Applications. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88974-717-7

Table of Contents

frontiers in Computer Science

EDITORIAL
published: 25 February 2022
doi: 10.3389/fcomp.2022.860619

Editorial: Brain-Computer Interfaces for Non-clinical (Home, Sports, Art, Entertainment, Education, Well-Being) Applications

Anton Nijholt[1], Jose Luis Contreras-Vidal[2], Camille Jeunet[3] and Aleksander Väljamäe[4]*

[1] Human Media Interaction, University of Twente, Enschede, Netherlands, [2] Industry/University Cooperative Research Center for Building Reliable Advances and Innovations in Neurotechnology, University of Houston, Houston, TX, United States, [3] Univ. Bordeaux, CNRS, EPHE, INCIA, UMR5287, Bordeaux, France, [4] School of Digital Technologies, Tallinn University, Tallinn, Estonia

Keywords: brain-computer interfaces, EEG, human-computer interaction, affective computing, non-clinical applications, fNIRS

Editorial on the Research Topic

Brain-Computer Interfaces for Non-clinical (Home, Sports, Art, Entertainment, Education, Well-Being) Applications

OPEN ACCESS

Edited and Reviewed by:
Egon L. Van Den Broek,
Utrecht University, Netherlands

****Correspondence:***
Anton Nijholt
a.nijholt@utwente.nl

Specialty section:
This article was submitted to Editorial Article, a section of the journal Frontiers in Computer Science

Received: *23 January 2022*
Accepted: *07 February 2022*
Published: *25 February 2022*

Citation:
Nijholt A, Contreras-Vidal JL, Jeunet C and Väljamäe A (2022) Editorial: Brain-Computer Interfaces for Non-clinical (Home, Sports, Art, Entertainment, Education, Well-Being) Applications. Front. Comput. Sci. 4:860619. doi: 10.3389/fcomp.2022.860619

INTRODUCTION

In this decade Brain-Computer Interface (BCI) technology has entered mainstream human-computer interaction (HCI) research for non-clinical applications. BCI has become part of multimodal interaction research as an additional interaction modality for a user of a technological system. BCI has also become part of research in which neurophysiological data provides a system with information about a user's affective and mental state, making it possible to adapt system, task, and interaction to a particular user, online (Fairclough, 2022). Currently, there is a market for inexpensive electroencephalographic (EEG) devices and software kits that capture voluntarily and involuntarily evoked brain activity and allow this activity to be translated into control and communication commands for environments and devices. Moreover, research on the use of deep networks for BCI applications has increased recently and promises to increase the accuracy of BCI systems (Craik et al., 2019). Overall, the availability of low cost non-invasive neurotechnology poses some ethical and regulatory challenges at the intersection of medical and consumer neurotechnologies.

Although EEG-based BCIs are limited in robustness and bandwidth, they are still, by far, the most accessible type of BCI to explore its potential use in domains such as games, entertainment, education, and art. While much of BCI research in clinics is increasingly relying on invasive recordings, such methods are most likely decades away from non-medical applications.

HCI researchers interest in BCI is increasing because the technology industry is expanding into application areas where efficiency is not the main goal of concern. Domestic or public space use of information and communication technology raise awareness of the importance of affect, comfort, family, community, or playfulness, rather than efficiency. Therefore, in addition to non-clinical BCI applications that require efficiency and precision, this Research Topic also addresses the use of BCI for various types of domestic, entertainment, educational, sports, and well-being applications. These applications can relate to an individual user as well as to multiple cooperating or competing users. We also see a renewed interest of artists to make use of such devices to design interactive art installations that know about the brain activity of an individual user or the collective brain activity of a group of users, for example, an audience (Contreras-Vidal et al., 2019). Hence, this Research Topic also addresses how BCI technology influences artistic creation and practice, and the use of BCI technology to manipulate and control sound, video, and virtual and augmented reality (VR/AR).

CONTRIBUTIONS

This Research Topic is composed of 11 accepted papers: seven dedicated to original research, a perspective, a mini review and two opinion pieces, and are dedicated to various themes and perspectives. These contributions address the multi-faceted nature of non-clinical BCIs, ranging from ethical ramifications of these neurotechnologies, applications to the arts, education, communication, wellbeing, and sports to the readiness of BCI deployment for gaming.

A first original research paper by Xie et al. focuses on the use of EEG-BCI technology as an *objective* evaluation method. The authors suggest using EEG and machine learning in order to assess car sound quality and show relationships between sound intensity and EEG energy.

Guillermo Bernal, Sean Montgomery and Pattie Maes discuss the future of BCI systems and how they may have an impact on productivity and efficiency, and/or augment the human experience by enhancing expressivity, understanding and empathy Bernal et al.. Artists and do-it-yoursef-ers (DIYers) are provided as examples of BCI users that deploy such systems beyond the intendedİ use (for additional discussion, the reader is referred to Paek et al., 2020). Current challenges on data and code: sharing, transparency (e.g., preprocessed data and closed-source algorithms typically used by commercial systems), lack of interpretability and explainability (e.g., black box systems), and security are recognized as barriers to democratizing BCI systems. Their manuscript concludes with a discussion of the impact of closed loop BCI systems for generative content and augmented cognition when coupled with VR/AR systems.

Currently, the P300 component of the event-related potential (ERP) certainly is one of the most reliable EEG patterns to be used to control a so-called reactive BCI. P300-BCIs are also relevant for fast stimulus recognition as both the amplitude and latency of this EEG component are linked to cognitive processes potentially triggered by those stimuli. In their original research paper, Sutaj et al. design and evaluate a novel P300-based real-time image ranking BCI. Their work demonstrates the relevance of BCIs for stimulus sorting and brings insights on the influence of stimulus-related cognitive processes on the associated EEG patterns.

Emotional reactions monitoring and classification is an important part of many real life applications involving physiological and affective computing technologies. Li et al. describe their open-source software toolbox MindLink-Eumpy that aims at recognizing emotions by integrating EEG and facial expression information. Besides online experiment conducted, the offline validation was done using DEAP (Koelstra et al., 2011) and MAHNOB-HCI (Soleymani et al., 2011) datasets. As expected, the results show that multimodal methods outperform single-modal methods in both offline and online experiments.

With the expansion of the BCI field comes the development of numerous applications. While they could be relevant and re-used, both for research purposes and out-of-the-lab applications (e.g., entertainment, rehabilitation, etc.), those applications most often remain confidential. Efforts to provide these application open-source, as in Woo et al., will be most valuable for the community.

Recently, functional near-infrared spectroscopy (fNIRS) has been for brain studies outside the lab or indoor conditions involving user physical activity like walking, as it is more robust to movement artifacts than EEG-based imaging. The study by de With et al. focused on fear of heights in VR exposure therapy (VRET) settings using head-mounted display. It focused on the question to what extent fNIRS can differentiate users with and without anxiety disorders. Two experimental groups involved controls and pre-screened individuals with fear of heights (based on Acrophobia Questionnaire). While results showed limited statistical significance, the experimental group showed stronger reactions to fear inducing stimuli. The study demonstrates ecological validity of combining fNIRS measurements and VRET, encouraging further work on BCI-based therapy applications.

Another fNIRS study by Slutter et al. addresses phenomenon of *choking under pressure* in soccer players just seconds before the actual penalty kick on the football field. Results showed that when experienced players were feeling anxious, their left temporal cortex activation increased. While the work was methodologically challenging (e.g., 59% of the collected data has been removed due to artifacts), it demonstrates an important trend of moving out-of-the-lab to get ecologically valid brain imaging results. In should be also noted, that due the study design and the sports topic, this work received high media coverage, with journalists often overinterpreting the original results.

Scott and Raftery's perspective contribution addresses the potential impact of integrating brain-computer interfaces (BCI) and creative art therapy approaches to promote rehabilitative and therapeutic interactions while increasing patient engagement Scott and Raftery. The authors rightly argue that the application of BCI systems informed by creative expression—as a new form of digital health tool—may be extended to promote emotional and physiological healing and recovery. However, it remains to be seen how artistic/creative BCIs can be personalized and prescribed for health and well-being for use individually or in social contexts (see recent collection of works on BCI for artistic expression in Nijholt, 2019).

In their mini-review, Joan Belo, Maureen Clerc and Daniele Schn discuss emergent applications of passive BCIs and EEG-based auditory attention detection (AAD) in education and art Belo et al.. Of interest is the proposed use of an AAD module, which monitors the source and location of the user's attended auditory sources, to guide BCI control, for example, of external sound generating instruments.

Affective brain-computer music interface technology is used for mood enhancement by providing users with suggestions for the music they like and modulating the users affective states. In an opinion piece by Hildt, social and ethical aspects of the technology are discussed with a focus on the role of the brain and considerations about responsibility for controlling one's affective state, neural profiling, and privacy-related aspects.

In the second opinion paper, Cattan discusses the readiness of BCI game interfaces for the general public. The paper has critical reflections on the common limitations discussed in the literature, such as low transfer rate, cost and encumberment of materials, and the lack of game design and graphics compared with video games available on the market. In his opinion, there should be more focus on qualitative aspects of the interaction with BCI games, where BCI should be limited to aspects that cannot be achieved by traditional inputs.

CONCLUSIONS

The picture emerging from the contributions to this Research Topic, with its wide range and different approaches, is one that highlights the emergent challenges and opportunities for non-clinical BCI systems and their potential impact on society. Clearly, this is an area of research that is gaining considerable attention and that is expected to reach the public sooner than clinical BCI systems, which require careful evaluation of risks and benefits to the end users, and approval by regulatory agencies. The reader is cautioned that deployment or modifications of non-clinical BCI systems by end-users may lead to unintended consequences that are currently poorly understood. This is particularly critical for closed-loop or neurofeedback systems that may alter or adapt cognitive-emotional states in healthy individuals or in person with mental or cognitive disabilities, or brain injury. Clearly more research is needed on the ethical and thrust-worthy application of BCI systems outside the clinic.

AUTHOR CONTRIBUTIONS

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

FUNDING

Partial support from the IUCRC BRAIN to JC-V is gladly recognized (NSF Award number 1650536). CJ has received partial support from the French National Agency (grant nos. ANR-20-CE38-0008-01—ABCIS project and ANR-20-CE37-0012-02—BETAPARK project). AV received partial support from the Estonian Research Council grant PUT1518.

ACKNOWLEDGMENTS

To the always helpful Frontiers team whose organizational skills and understanding made this Research Topic possible.

REFERENCES

Contreras-Vidal, J. L., Robleto, D., Cruz-Garza, J. G., Azorín, J. M., and Nam, C. S. (2019). *Mobile Brain-Body Imaging and the Neuroscience of Art, Innovation and Creativity*. Cham: Springer.

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

Fairclough, S. H. (2022). "Designing human-computer interaction with neuroadaptive technology," in *Current Research in Neuroadaptive Technology* (London: Academic Press), 1–15.

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., et al. (2011). Deap: a database for emotion analysis; using physiological signals. *IEEE Trans. Affect. Comput.* 3, 18–31. doi: 10.1109/T-AFFC.2011.15

Nijholt, A. (2019). *Brain Art: Brain-computer Interfaces for Artistic Expression*. Cham: Springer.

Paek, A. Y., Brantley, J. A., Evans, B. J., and Contreras-Vidal, J. L. (2020). Concerns in the blurred divisions between medical and consumer neurotechnology. *IEEE Syst. J.* 15, 3069–3080. doi: 10.1109/JSYST.2020.3032609

Soleymani, M., Lichtenauer, J., Pun, T., and Pantic, M. (2011). A multimodal database for affect recognition and implicit tagging. *IEEE Trans. Affect. Comput.* 3, 42–55. doi: 10.1109/T-AFFC.2011.25

Conflict of Interest: JC-V is the Site Director for the IUCRC BRAIN at the University of Houston-an industry-university cooperative research center that is funded by the National Science Foundation and industrial partners.

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

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 19 February 2021
doi: 10.3389/fnhum.2021.621493

MindLink-Eumpy: An Open-Source Python Toolbox for Multimodal Emotion Recognition

Ruixin Li[1†], Yan Liang[1†], Xiaojian Liu[1], Bingbing Wang[1], Wenxin Huang[1], Zhaoxin Cai[1], Yaoguang Ye[1], Lina Qiu[1] and Jiahui Pan[1,2]*

[1] School of Software, South China Normal University, Guangzhou, China, [2] Pazhou Lab, Guangzhou, China

Emotion recognition plays an important role in intelligent human–computer interaction, but the related research still faces the problems of low accuracy and subject dependence. In this paper, an open-source software toolbox called MindLink-Eumpy is developed to recognize emotions by integrating electroencephalogram (EEG) and facial expression information. MindLink-Eumpy first applies a series of tools to automatically obtain physiological data from subjects and then analyzes the obtained facial expression data and EEG data, respectively, and finally fuses the two different signals at a decision level. In the detection of facial expressions, the algorithm used by MindLink-Eumpy is a multitask convolutional neural network (CNN) based on transfer learning technique. In the detection of EEG, MindLink-Eumpy provides two algorithms, including a subject-dependent model based on support vector machine (SVM) and a subject-independent model based on long short-term memory network (LSTM). In the decision-level fusion, weight enumerator and AdaBoost technique are applied to combine the predictions of SVM and CNN. We conducted two offline experiments on the Database for Emotion Analysis Using Physiological Signals (DEAP) dataset and the Multimodal Database for Affect Recognition and Implicit Tagging (MAHNOB-HCI) dataset, respectively, and conducted an online experiment on 15 healthy subjects. The results show that multimodal methods outperform single-modal methods in both offline and online experiments. In the subject-dependent condition, the multimodal method achieved an accuracy of 71.00% in the valence dimension and an accuracy of 72.14% in the arousal dimension. In the subject-independent condition, the LSTM-based method achieved an accuracy of 78.56% in the valence dimension and an accuracy of 77.22% in the arousal dimension. The feasibility and efficiency of MindLink-Eumpy for emotion recognition is thus demonstrated.

Keywords: multimodal emotion recognition, multitask convolutional neural network (CNN), support vector machine (SVM), subject-independent method, long short-term memory network (LSTM)

OPEN ACCESS

Edited by:
Anton Nijholt,
University of Twente, Netherlands

Reviewed by:
Pietro Aricò,
Sapienza University of Rome, Italy
Yisi Liu,
Fraunhofer Singapore, Singapore

***Correspondence:**
Jiahui Pan
panjiahui@m.scnu.edu.cn;
panjh82@qq.com

† 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: *26 October 2020*
Accepted: *18 January 2021*
Published: *19 February 2021*

Citation:
Li R, Liang Y, Liu X, Wang B, Huang W, Cai Z, Ye Y, Qiu L and Pan J (2021) MindLink-Eumpy: An Open-Source Python Toolbox for Multimodal Emotion Recognition. Front. Hum. Neurosci. 15:621493. doi: 10.3389/fnhum.2021.621493

INTRODUCTION

Emotions are biological states associated with the nervous system (Damasio, 1998), and its changes are related to subjective feelings and objective behavioral responses. Emotion recognition plays an essential role in human–computer interaction. It is an emerging interdisciplinary research field that covers various methods and techniques in artificial intelligence (AI), natural language processing

(NLP), and cognition and social sciences (Poria et al., 2017). Although the studies on emotion recognition have made great improvements in recent years, there are still limitations such as low accuracy and subject dependence. Thus, there is an urgent need for an innovative toolbox with effective methods to enlarge the dataset and improve the accuracy of emotion recognition.

Previous studies (Davidson et al., 1990) exerted the Approach/Withdrawal index as an emotional indicator of the relationship between emotion, approach, and withdrawal. Gianluca et al. (Di Flumeri et al., 2017) demonstrated the reliability of this index of pleasantness. Afterward, some scholars found that algorithms based on transfer learning, fusion of multimodal information, or subject-independent methods can improve the performance of emotion recognition. For example, Nguyen et al. (2018) proposed a novel transfer learning approach based on PathNet and conduct various experiments on the Surrey Audio-Visual Expressed Emotion (SAVEE) dataset and the eNTERFACE dataset and found that this approach could improve the performance of emotion recognition. Sebe et al. (2005) conducted a survey and pointed out that multimodal emotion recognition (such as the combination of facial information, voice, and physiological signals) achieved higher accuracy than traditional single-modal emotion recognition. Furthermore, Georgieva et al. (2015) compared six unsupervised machine learning methods and performed experiments for intersubject models and intrasubject models. The results showed that event-related potential (ERP) clustering (especially the Fuzzy C-means clustering) algorithm was a promising approach that can extract statistical underlying correlations of brain activity among subjects to decode the human emotional state. However, all the above studies did not combine the electroencephalogram (EEG) modality using deep learning technology in subject-independent emotion recognition.

Furthermore, the influence of emotion is manifested in a variety of levels and modalities. On the one hand, peripheral signals (such as facial expressions, verbal speech, and body language) are related to the somatic nervous system and can reflect changes in emotion states. On the other hand, many studies often assessed the power spectra of EEG in different frequency bands to examine their relationship with emotional states. For example, literature (Alsolamy and Fattouh, 2016; Guzel Aydin et al., 2016; Jiang et al., 2019) reported several spectral changes and brain regions related to emotional responses, such as the changes of theta (θ: 4–7 Hz) power in the right parietal lobe, the asymmetry of alpha (α: 8–13 Hz) power in the anterior area of the brain, the asymmetry of beta (β: 14–30 Hz) power in the parietal region, and the changes of gamma (γ: 31–50 Hz) power in the right parietal regions. Most of the previous studies have used peripheral signals or brain signals alone to identify emotions, and little attention has been paid to the fusion between the brain and peripheral signals.

To overcome the abovementioned difficulties in emotion recognition, we proposed an open-source free toolbox named MindLink-Eumpy. MindLink-Eumpy mainly focuses on the recognition of continuous movie-induced emotions rather than discrete emotions (Yan den Broek, 2013) (the basic discrete emotions include happiness, sadness, surprise, fear, anger, and disgust) and provides a series of tools for physiological data processing. MindLink-Eumpy applies vector models to classify emotions because they can quantify emotions better than circumplex models (a concept of discrete emotions) (Posner et al., 2005). Furthermore, MindLink-Eumpy adopts a valence–arousal model (Lang et al., 1997) (continuous emotion) and self-assessment manikins (SAMs) (Bradley and Lang, 1994) to evaluate emotion. The scores of valence and arousal dimensions are both between 1 and 9 (all scores are integers). The valence reflects the level of pleasure, and arousal reflects the level of intensity. High scores represent high levels of pleasure or intensity.

MindLink-Eumpy provides a series of continuous emotion recognition methods based on facial expressions and EEG signals. Specifically, as a toolbox designed for scientific research, MindLink-Eumpy is suitable not only for emotion recognition based on the public databases Database for Emotion Analysis Using Physiological Signals (DEAP) (Koelstra et al., 2012) and MAHNOB-HCI (Soleymani et al., 2012) but also for emotion recognition based on self-created databases. To acquire physiological data and create our own database, MindLink-Eumpy provides recorders (programs for device control, especially programs for collecting data) to collect EEG signals and facial images. Moreover, MindLink-Eumpy implements an event-related potential (ERP) (Sur and Sinha, 2009) paradigm in a Web-based framework to induce the subject's emotions through video clips. In the detection of facial expression, MindLink-Eumpy uses multitask convolutional neural networks (CNNs) (Lawrence et al., 1997) based on transfer learning techniques to overcome the common problem of lack of data and achieve higher accuracy. MindLink-Eumpy offers two methods in the detection of EEG. One is a subject-dependent model based on support vector machine (SVM) (Cortes and Vapnik, 1995), which is able to achieve high accuracy when the validation data and the training data are homogeneous. The other one is a subject-independent model based on long short-term memory network (LSTM) (Hochreiter and Schmidhuber, 1997; Koelstra et al., 2012), which is used to reduce the effects caused by the individual variations and non-stationarity of EEG signals. The latter method yields more stable performance when the validation data and training data are heterogeneous. Moreover, to improve the accuracy of emotion recognition for homogeneous data, MindLink-Eumpy proposes two decision-level fusion methods for multimodal emotion recognition tasks, namely, weight enumerator and adaptive boosting (AdaBoost) technique (Das et al., 2015), to fuse the decision-level information of SVM and CNN. For the heterogeneous data, the subject-independent method we used is the EEG-based LSTM model. Our experimental results show that when the validation data and training data are homogeneous, the highest average accuracy achieved by the multimodal subject-dependent models in the arousal and valence dimensions were 72.14% and 71.00%, respectively. However, when the validation data and training data are heterogeneous, the highest average accuracy achieved by the EEG-based subject-independent model in the arousal and valence dimensions were 77.22% and 78.56%, respectively.

This paper introduces an open-source Python toolbox for multimodal emotion recognition, MindLink-Eumpy, including its structure, related algorithms, and functions. The *Introduction* section of this paper covers the background and significance of this work. The *Related Work* section introduces the related works on emotion recognition and some related toolboxes. The *MindLink-Eumpy: Architecture, Modules, and Features* section presents MindLink-Eumpy and describes its structure, methods, and functions in detail. The *Methods for Emotion Recognition* section describes the continuous emotion recognition methods used in MindLink-Eumpy in detail and proposes multimodal subject-dependent methods and EEG-based subject-independent methods. The *Experiments and Results* section demonstrates the innovations and effectiveness of MindLink-Eumpy. The *Discussion and Conclusion* section summarizes the advantages and limitations of MindLink-Eumpy compared with other state-of-the-art methods, as well as its potential applications and areas of future work.

RELATED WORK

In this section, we briefly review some related work on emotion recognition. This section includes three subsections: **(i)** related studies on emotion recognition, **(ii)** software-related emotion recognition, and **(iii)** comparison of related software with MindLink-Eumpy.

Related Research on Emotion Recognition

It is well known that emotion recognition techniques have yielded considerable improvements in the past few years, and here come some articles that inspired us during our study. To begin with, transfer learning technique has the potential to tackle the difficulties of small datasets in emotion recognition area. Facial expression recognition needs plenty of facial images but it is hard to recruit enough subjects. To deal with this problem, Prakash et al. (2019) proposed an automatic facial emotion recognition method using CNNs with a transfer learning approach. This approach was demonstrated to be effective with an average accuracy over 98% in their experiments. Furthermore, we tried to combine facial expression and EEG because the performance of multimodal emotion recognition methods is superior to that of single-modal methods, which has been demonstrated long before. As an example, in a research in 2008: Kessous et al. (Castellano et al., 2008) integrated information from facial expressions, body movements, gestures, and speech and found that the multimodal approach improved accuracy by more than 10% compared to the most successful single-modal system. Finally, we tried to improve subject independence of EEG-based methods because MindLink-Eumpy attaches more importance to human neuroscience. The subject-independent emotion recognition based on EEG signals is the current research hotspot. Under this circumstance, Alhagry et al. (2017) proposed an end-to-end LSTM-recurrent neural network (RNN) to analyze emotion from raw EEG signals, in which they achieved average accuracy rates of 85.65%, 85.45%, and 87.99% in classification for the arousal, valence, and fondness dimensions, respectively. Therefore, we theoretically chose LSTM as the first subject-independent method of MindLink-Eumpy.

Software Toolboxes Related to Emotion Recognition

This subsection briefly introduces three software toolboxes for emotion recognition that are currently used in both scientific research and industrial applications.

Computer Expression Recognition Toolbox (CERT)

The Computer Expression Recognition Toolbox (CERT) (Littlewort et al., 2011) is an open-source free software tool for fully automatic real-time facial expression recognition. It can automatically code the intensity of 19 different facial actions from the Facial Action Unit Coding System (FACS) and six different prototypical facial expressions. Moreover, this tool can estimate the positions of 10 facial features and the 3D orientation (yaw, pitch, and roll) of the head. Previous experiments have demonstrated that CERT can achieve an accuracy of nearly 80% when applied to a spontaneous facial expression dataset (Littlewort et al., 2011).

MixedEmotions

The MixedEmotions toolbox (Buitelaar et al., 2018) contains text, audio, and video processing functions aimed at emotion recognition and provides a plug-and-play and ready-to-use set of emotion recognition modules. The current version is mainly applied to three real-world cases: emotion-driven smart TV use (emotion-based recommendation), brand reputation analysis (monitoring the reputation of a brand from tweets and YouTube videos), and call center monitoring (monitoring the emotions of customers in a help desk setting).

Toolbox for Emotional Feature Extraction From Physiological Signals (TEAP)

The Toolbox for Emotional Feature Extraction from Physiological Signals (TEAP) (Soleymani et al., 2017) is an open-source MATLAB toolbox that can process and calculate emotion-related features from multiple physiological signals, including EEG, galvanic skin response (GSR), electromyogram (EMG), skin temperature, respiration pattern, and blood volume pulse information. The toolbox has been tested on the MAHNOB-HCI and DEAP databases and has shown promising performance (Soleymani et al., 2017).

Comparison With MindLink-Eumpy

This subsection compares the differences between software toolboxes for emotion recognition and describes the advantages of MindLink-Eumpy. **Table 1** lists the programming languages and functions of the above toolboxes.

MindLink-Eumpy is an open-source Python toolbox with modular tools and frameworks for different functions. The main functions are **(i)** providing a framework for online ERP experiments, **(ii)** reading real-time data from devices during online experiments and practical usage scenario, **(iii)** processing multimodal data including facial images and EEG signals, **(iv)** providing model training interfaces and datasets storage medium

TABLE 1 | Comparison of different toolboxes related with emotion recognition.

Toolbox Name	Programming Language	Main Features
CERT (Littlewort et al., 2011)	Python	- Fully automatic facial expression recognition in real time - Automatically encodes the intensity of 19 different facial actions from FACS and estimates the positions of facial features and the 3D orientation of the head
MixedEmotions (Buitelaar et al., 2018)	Python	- Provides a plug-and-play and ready-to-use set of emotion recognition modules - Provides a unified solution for large-scale emotion analysis on heterogeneous, multimodal, text, speech, video, and social media data streams
TEAP (Soleymani et al., 2017)	MATLAB, Octave	- Imports, processes, and visualizes physiological signals - Processes and calculates emotionally relevant features
MindLink-Eumpy	Python	- Two approaches of facial expression and EEG for emotion recognition - Two decision-level fusion methods for fusion of sub-classifiers in different modalities to improve accuracy - Reads, processes, visualizes multimodal real-time data (facial images and EEG signals) and stores data into a folder system - Subject-independent emotion recognition approach based on LSTM in EEG modality

CERT, Computer expression recognition toolbox; FACS, Facial action unit coding system.

(here we call it a database), and **(v)** real-time emotion recognition and data visualization.

By combining decision-level information of facial expressions and EEG, MindLink-Eumpy has obtained a promising accuracy of emotion recognition. Moreover, MindLink-Eumpy provides an LSTM model based on EEG for subject-independent emotion recognition. The abovementioned toolboxes represent the state-of-the-art software in emotion recognition area. While most of the toolboxes provide tools for data processing or methods for emotion recognition, seldom do they focus on the scientific research on human neuroscience and practical applications. MindLink-Eumpy provides tools for EEG collection, preprocessing, and display so as to reflect emotions straightforward. To enhance stability in practical application, MindLink-Eumpy provides tools for facial images, including functions of images processing based on OpenCV and a CNN model for emotion recognition.

MINDLINK-EUMPY: ARCHITECTURE, MODULES, AND FEATURES

This section gives an overview about MindLink-Eumpy, including **(i)** the architecture of MindLink-Eumpy shown in **Figure 1**, **(ii)** the modules in MindLink-Eumpy, and **(iii)** the features of MindLink-Eumpy. In real-time running, fusion tools are used in the step of fused scores.

Architecture of MindLink-Eumpy

MindLink-Eumpy can be mainly separated into two parts: database creation (online experiment paradigm) and real-time detection framework. In the modules of database creation, MindLink-Eumpy provides a series of tools for data streaming and decoding from devices (brain–computer interfaces and cameras). During the online experiment, data from the subject will be stored into a folder system (here, we call it a database or a data storage medium). In real-time detection framework, MindLink-Eumpy provides visualization function for EEG, facial images, and analyses of emotion.

Specifically, MindLink-Eumpy utilizes existing Python open-source libraries such as Numpy (van der Walt et al., 2011), scikit-learn, TensorFlow, Keras (Pedregosa et al., 2011), Flask, Pandas, and others. In the database module, we adopted two public emotion databases, i.e., DEAP and MAHNOB-HCI, and created our own database to evaluate the performance of the proposed methods. Emotions are elicited by video clips from commercial films. In the experimental paradigm module, videos in the database are applied to elicit emotion, during which MindLink-Eumpy can use the readers (an EEG reader in the EEG toolbox and a facial image reader in the facial images toolbox) to read and save physiological data. The details of the experiments and evaluation are presented in the *Experiments and Results* section. In the algorithm modules, we integrated all methods into three packages: facial images tools, EEG tools, and fusion tools. In addition to data processing methods, MindLink-Eumpy also provides emotion classification methods, including the multitask CNN method in the facial image toolbox, the SVM and an LSTM in the EEG toolbox, and two methods in the fusion toolbox for decision-level fusion. More details of these methods are shown in *Methods for Emotion Recognition* section.

Modules of MindLink-Eumpy

Database Creation

To address the lack of data problem, we designed a data collection framework to acquire and store facial images and EEG signals. By this module, MindLink-Eumpy can help conduct online experiments and obtain subjects' data more simply. New methods can be validated more effectively in public databases and our own database. One disadvantage about public databases is that subjects' emotion and related feedback may differ from culture, gender, and other uncontrollable factors. Therefore, our research lacks data from subjects similar to actual users of MindLink-Eumpy. The function of database creation aims at eliminating the problem of low accuracy of practical application caused by domain differences of data.

Device invocations

Physiological data streams are first recorded by hardware such as EEG acquisition equipment (e.g., Emotiv EPOC+ headset in this

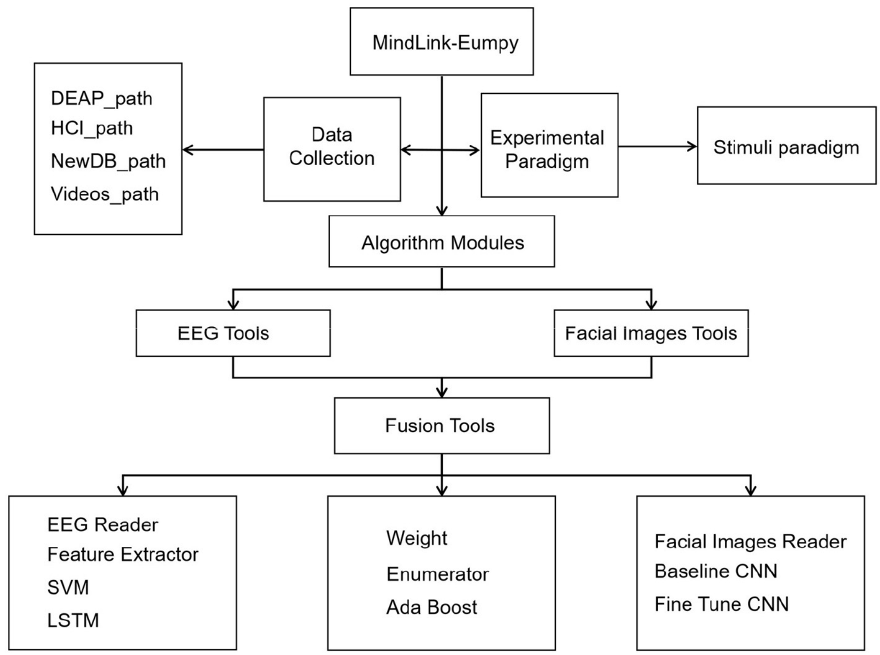


FIGURE 1 | Architecture of MindLink-Eumpy.

study) and optical camera. Then, the hardware sends data streams to the back end of MindLink-Eumpy. This kind of process is called data streaming, and we designed programs called readers to conduct the data streaming process. MindLink-Eumpy has two readers (an EEG reader and a facial image reader) to obtain EEG signals and facial images. The EEG reader invokes the driver of the Emotiv EPOC+ headset suitable for MindLink-Eumpy and uses the interfaces of the corresponding software development kit (SDK) to obtain EEG data and store it in the computer memory. The main process of facial image reading is the same as the reading process of EEG signals, but the facial image reader uses the OpenCV library to start the camera and obtain digital images stored in the computer memory.

Data storage

MindLink-Eumpy provides a prototype storage medium for data relating to emotions. In the computer memory, MindLink-Eumpy establishes a queue for temporary data storage. By controlling the size of this queue, MindLink-Eumpy is able to synchronize the frequency of data refreshing by devices and the Python-Flask back end to prevent data explosion. However, in external memory, from the perspective of the file format, raw EEG signals are required to be stored in (*.fif) files, but power spectral density (PSD) data (Ng et al., 2019) of the EEG signals are stored as matrices in (*.npy) files through the Numpy library. For facial images, MindLink-Eumpy uses OpenCV to save temporary images as videos in (*.mp4) files. Ground-truth labels and the personal information of subjects are saved in (*.csv) files through the Pandas library. We can access databases with a string of the subject's information (reported before an experimental trial starts).

Real-Time Detection Framework

We designed a real-time detection framework to widen the scope of application of MindLink-Eumpy. Based on data readers, the Python-Flask back end, and web technology, this framework applies the E-Charts technique to visualize real-time data.

Electroencephalogram detection

In EEG detection, MindLink-Eumpy first modifies the real-time EEG data temporarily stored in the computer memory into a specific format and sends to the front end. Then, the EEG data are displayed on a web page with a visually appealing style. The lower panel of **Figure 2** shows the EEG signals of five channels (AF3, AF4, T7, T8, and Pz) in a two-dimensional coordinate system, and the upper panel shows the mapping of PSD data (theta, alpha, beta, and gamma) on brain patterns to reflect the effects of valence and arousal levels in different brain regions. In graphical user interface (GUI), different colors represent different brain regions, and the brightness of the color represents the value of the PSD. Specifically, yellow represents the region where AF3 and AF4 are located, red represents the region where T7 and T8 are located, and purple represents the area where Pz is located. In this study, five channels (AF3, AF4, T7, T8, and Pz) were selected for the default display. Users can also manually select other channels based on their equipment. Furthermore, PSD is the most commonly used feature in emotion recognition (Park et al., 2013). Thus, in the current version of MindLink-Eumpy, we only provide the function of PSD pattern.

Facial expression detection

In real-time facial image detection, MindLink-Eumpy first uses the Viola–Jones face detector (Viola and Jones, 2001) to detect

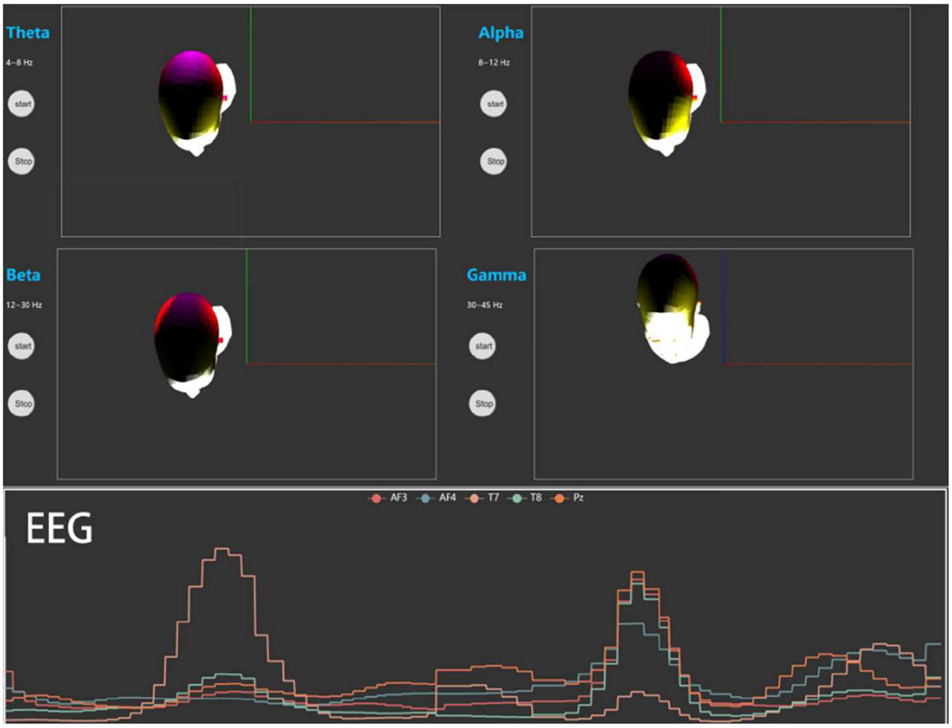


FIGURE 2 | Visualization for power spectral density (PSD) and electroencephalogram (EEG).

the face of the subject. Then, facial features are identified by the multitask CNN. **Figure 3** shows the calculation results of three layers in the CNN during a forward pass on a web page. For the first and second convolutional layers, the low-level features such as edges and light are displayed. For the final convolutional layer, the high-level features such as the eyes and mouth of a user are displayed.

Emotion visualization

Figure 4 shows the main screen of the MindLink-Eumpy's GUI, which provides visualization functions for EEG data, facial images, continuous emotions (valence–arousal emotion model), and discrete emotions. Continuous emotions can be obtained by fusion methods that combine the predictions of the SVM and CNN in decision level. In this study, the K-nearest neighbors (KNN) was used to transform the continuous emotions to 16 discrete emotions, including pride, elation, joy, satisfaction, relief, hope, interest, surprise, sadness, fear, shame, guilt, envy, disgust, contempt, and anger. Specifically, 16 samples with ground-truth labels in the dataset were first set, and then 16 categories were classified according to Euclidean distance. This function is designed to intuitively display emotions in GUI for users. The intensity of emotions is plotted on a radar map (emotion wheel). In MindLink-Eumpy, both continuous and discrete emotions are sent from the back end to the front end and are displayed in real time. Visual emotion data are displayed in the upper right of the screen. We can click a white button in the middle of the screen to switch interfaces between continuous emotion and discrete emotion.

Features of MindLink-Eumpy

Herein, we summarize the features of our toolbox. Notably, the toolbox is characterized by simple data acquisition and storage, high accuracy based on multimodal emotion recognition, low algorithmic complexity, and subject independence. We have established a framework for academic experiments, and MindLink-Eumpy provides tools for data streaming, processing, and storage. Moreover, MindLink-Eumpy provides machine learning algorithms and deep learning techniques with promising performance based on both accuracy and algorithm complexity. These accurate methods are based on multimodal emotion recognition, and subject independence is achieved with single-modal EEG data. Furthermore, users can save and access collected data and models that have been trained to customize databases and methods. In short, our toolbox incorporates database, experimental paradigm, and software tools, which allow developers to easily extend model functionality and optimize usability in collaborative development.

METHODS FOR EMOTION RECOGNITION

Workflows of the Emotion Recognition Methods

This section describes the multimodal emotion recognition methods of subject dependence, which are provided in the algorithm modules of MindLink-Eumpy, as shown in **Figure 5**.

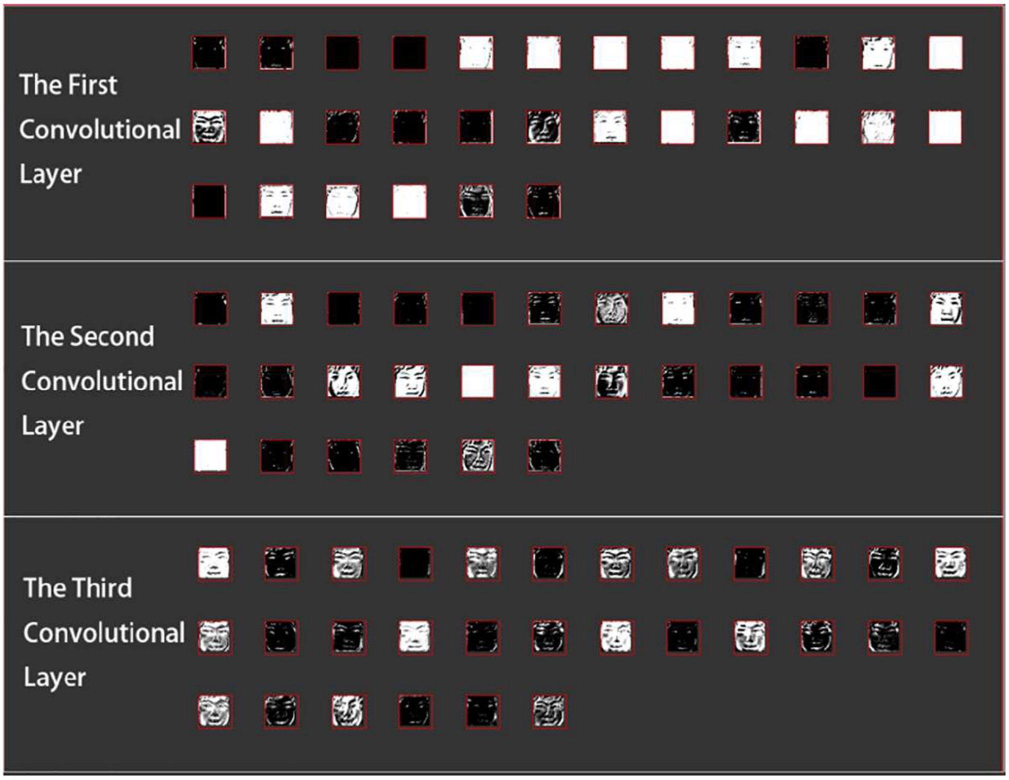


FIGURE 3 | Visualization of facial detection and feature identification by a convolutional neural network (CNN).

In the facial image tools submodule, the baseline CNN was pre-trained with a large open-source database, and the CNN modified with a database generated by the authors was a well-trained multitask model. Next, in the EEG tools submodule, the features of data were extracted with a feature extractor and were fed to the SVM and LSTM network. Finally, in the fusion tools submodule, the weight enumerator and AdaBoost method were applied to combine the predictions of the SVM (EEG modality) and CNN (facial expression modality) in decision level. It is worth mentioning that the combination of the SVM and CNN is subject dependent. Therefore, in the experiments, we trained one particular model for each subject separately.

Facial Expression Detection

Architecture of Multitask Convolutional Neural Network

We used a kind of transfer learning technique with a public database and our database to train the multitask CNN model for feature extraction and emotion classification. To obtain a well-trained multitask CNN, first, we pre-trained the CNN using a large regular database FER-2013 with image-level annotations (Goodfellow et al., 2015). Second, we froze all the parameters of the baseline CNN (the first three convolutional layers) and conducted a stochastic gradient descent (SGD) training (fine-tuning) by using a specific small database, while setting the learning rate to 0.0001.

In real-time detection, the images extracted from a video were input into the well-trained multitask CNN, so that we obtained multiple sets of valence and arousal scores. The highest scores were chosen to be the final valence and arousal scores. In addition, we resampled the videos to 4 Hz and used OpenCV to obtain grayscale images (640 × 480 PNG). Then, the Viola–Jones face detector was used to find the facial position in the image frame.

The size of the input images was 48 × 48 × 1 (grayscale images). A dropout layer with a deactivation rate of 0.5 was applied between the output and the dense layer to partially mitigate overfitting. The first, second, and third layers were convolution layers, and the fourth layer was a fully connected layer. The first layer has 32 convolution kernels with a size of 3 × 3 × 1. We used padding for the first convolutional layer. Padding is the addition of null pixels to increase the size of an image. Null pixels here refer to pixels with a value of 0. We used Keras to implement padding and CNN. Here, we have a 48 × 48 × 1 image and a 3 × 3 × 1 filter. With padding, the size of the first input image could be enlarged to 50 × 50 × 1, and the output of the convolutional layer (the second layer) could be 48 × 48 × 1, which preserves the same size as the original input image. The second layer had 32 convolution kernels with a size of 3 × 3 × 32. The third layer had 64 convolution kernels with a size of 3 × 3 × 32. The fourth layer was fully connected to 64 neurons. The final output layer had output valence and arousal scores for given emotion states.

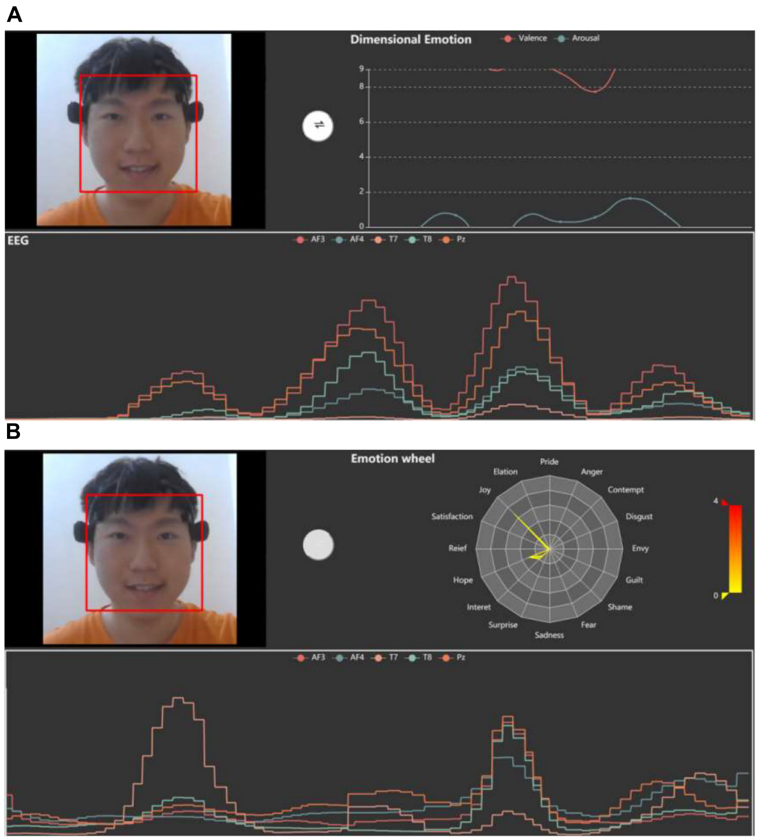


FIGURE 4 | Visualizations for continuous emotion and discrete emotion. **(A)** Graphical user interface (GUI) for continuous emotion. **(B)** GUI for discrete emotion.

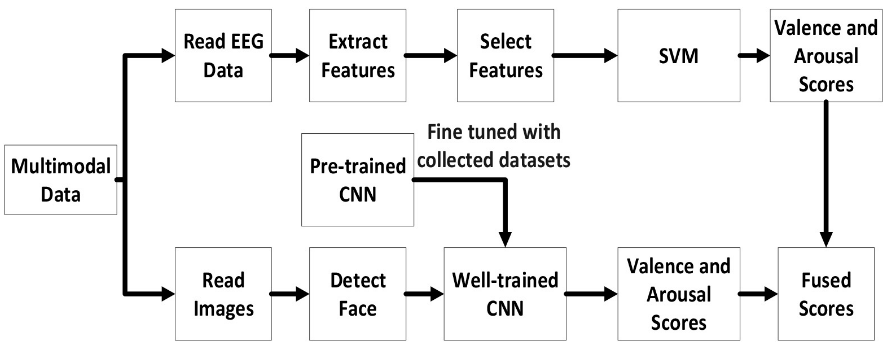


FIGURE 5 | Workflow of the subject-dependent approach in MindLink-Eumpy.

All convolutional layers and fully connected layers included a rectified linear unit (***ReLU***) as the activation function. Finally, the multitask CNN included two fully connected layers for separating valence and arousal scores. **Figure 6** shows the architecture of the multitask CNN.

Emotion Computing Based on Facial Expression

The first branch of the fully connected layer was used to calculate the valence scores, and the second branch was used to calculate the arousal scores. The output scores were sent to a ***sigmoid*** function to minimize the cross-entropy loss. Equation (1) represents the loss functionL_n.

$$L_n = -\sum_{i=1}^{m} \left(1 - y_{ni} \log \hat{y}_{ni}\right) \log\left(1 - \hat{y}_{ni}\right) \tag{1}$$

In equation (1), $\boldsymbol{n}$ represents the branch of the fully connected layer (when $\boldsymbol{n}$ is 1, L_n is the loss function of the valence branch; and when $\boldsymbol{n}$ is 2, L_n is the loss function of the arousal branch), y_{ni}represents the ground-truth labels for the $_{ith}$ sample,

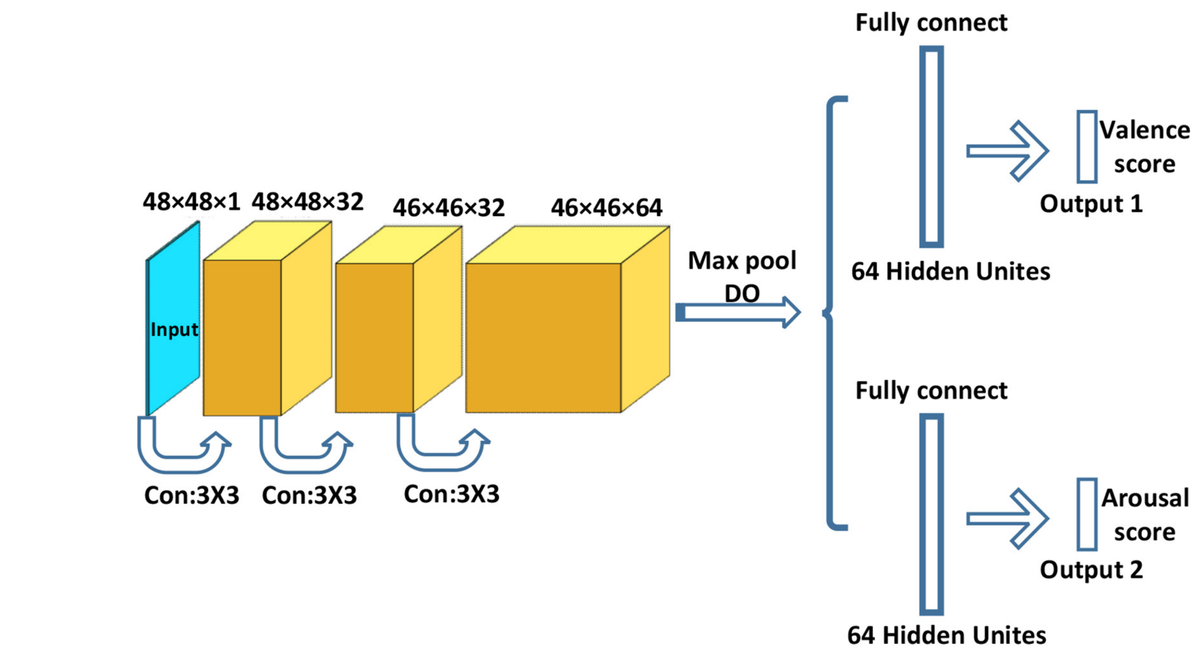


FIGURE 6 | Multitask convolutional neural network. "DO" represents the dropout layer.

$\hat{y}_{ni}$ represents the ***sigmoid*** outputs for the $_{ith}$ sample, and $\boldsymbol{m}$ represents the size of the training sample.

Finally, we minimized the linear combination of L_1 and L_2. Equation (2) represents the linear combination of the loss functions.

$$L = \sum_{i=1}^{2} \alpha_i L_i \tag{2}$$

In equation (2), $\alpha_i$represents linear weights. Notably, if any of α_i is set to 0, the model returns to a single-task CNN model.

After emotion regression calculations, we classified emotions based on equation (3).

$$r_{face} = \begin{cases} high & S_{face} \geq 0.5 \\ low & S_{face} < 0.5 \end{cases} \tag{3}$$

In equation (3), r_{face}represents the classification results of the valence or arousal dimension associated with a facial expression, and S_{face} represents the scores of regression calculations. Thus, the valence and arousal dimension scores are dichotomized at high and low levels, and there are a total of four emotion categories: high valence and high arousal, high valence and low arousal, low valence and high arousal, low valence and low arousal.

Electroencephalogram-Based Emotion Recognition

This subsection describes the emotion recognition methods in MindLink-Eumpy based on EEG signals. Here, we introduce the workflow of EEG-based emotion recognition. First, we used the Emotiv EPOC+ headset to record online EEG signals. Then, the multitaper method with fast Fourier transform (FFT) (Thomson, 1982) was used to extract the PSD features of the EEG signals. Finally, all the features were input into the SVM or LSTM for emotion recognition.

Electroencephalogram Data Processing and Feature Extraction

Previous studies (Alsolamy and Fattouh, 2016; Guzel Aydin et al., 2016; Jiang et al., 2019) have demonstrated that the PSDs of theta, alpha, beta, and gamma in the frontal, temporal, and occipital regions of the brain are highly related to human emotion. In this study, the features of EEG were extracted and selected based on these findings. In order to reduce the occurrence of artifacts, we first issued proper instructions to the subjects and repeatedly instructed the subjects to avoid blinking or moving their bodies during the experiment. Then, the filtering and multitaper (Thomson, 1982) techniques were used to remove the artifacts and keep the related neurological phenomenon intact in data processing.

It is worth noting that, in this study, we proposed two different EEG processing approaches for the online analysis and the offline analysis. Specifically, in the online analysis, we first remove artifacts through the function provided by the software development kit (SDK) of the Emotiv EPOC+ headset. Then, we used FFT to calculate PSDs, which is also provided by this equipment. It should be stressed that our EEG recording equipment Emotiv EPOC+ headset is designed for emotion recognition, and it can capture the EEG data from the following 14 channels located in the frontal, temporal, and occipital lobes: AF3, F3, F7, FC5, T7, P7, O1, AF4, F4, F8, FC6, T8, P8, and O2. However, in the offline analysis, the EEG data were bandpass filtered on five frequency bands (theta, slow alpha, alpha, beta, and gamma) from 4 to 45 Hz by finite impulse response (FIR) filters, and then the corresponding PSDs were obtained by FFT (overlap: 50%, time window: 1 s). In order to improve the accuracy in offline analysis of public databases, we selected 14 channels different from the online analysis: Fp1, T7, CP1, Oz, Fp2, F8, FC6, FC2, Cz, C4, T8, CP6, CP2, PO4. Finally, the PSDs of 14 channels and three symmetric pairs of channels ("T7–T8," "Fp1–Fp2," and "CP1–CP2") were used as the EEG features.

Subject-Dependent Method Based on Support Vector Machine

After reading EEG signals from the above electrodes, five PSD features were extracted: theta (4 Hz < f < 8 Hz), slow alpha (8 Hz < f < 10 Hz), alpha (10 Hz < f < 12 Hz), beta (12 Hz < f < 30 Hz), and gamma (30 Hz < f < 45 Hz) features. The total number of EEG features was 70 (14 × 5 = 70). MindLink-Eumpy uses SVM-RFE (recursive feature elimination) to select optimal features by iteratively calculating feature weights of the linear SVM classifier and subsequently removing the 10% features with the lowest weights. Then, we split the selected features with 10-fold inner cross-validation for the training set (Duan et al., 2005). Following the facial expression modality tasks, we classified emotion according to equation (4).

$$r_{EEG} = \begin{cases} high & S_{EEG} \geq 0.5 \\ low & S_{EEG} < 0.5 \end{cases} \tag{4}$$

In equation (4), r_{EEG}represents the classification result for the valence or arousal dimension in the EEG modality. S_{EEG}represents the scores of the regression calculations in the EEG modality.

Subject-Independent Method Based on Long Short-Term Memory

In this paper, we proposed an EEG-based subject-independent emotion recognition method based on LSTM. MindLink-Eumpy provides a well-trained model for this method. To implement this method, we constructed the features of the time sequence and performed regression calculations based on the LSTM. Since we applied supervised learning techniques to train the model, the data of the subjects in the experiment are all labeled and have the same distribution, resulting in the high accuracy of the experimental results. However, most of the subject-independent methods applied semi-supervised learning techniques or transfer learning techniques (Rodrigues et al., 2019; Li et al., 2020). This means that due to the domain differences between different subjects, the data in the model evaluation process cannot be labeled and its data distribution is different from the data in the model training process, which leads to a degradation in performance.

Here, we used the wavelet transform algorithm described above to extract features. Additionally, we regarded data every 10 s as a set of samples and sampled the data with an overlap rate of 50%. In the offline experiments using the MAHNOB-HCI database, 85 features were sampled per second. We picked 14 channels from the MAHNOB-HCI database and three symmetrical channel pairs of EEG data. For each channel, five PSD features are extracted from raw EEG data. Thus, the number of EEG features is (14 + 3) × 5 = 85. Each 10-s sample was used to construct a matrix with the size of 10 × 85, of which the first dimension is 10 and the second is 85. In this way, we avoided temporarily saving features in a one-dimensional vector unsuitable for the LSTM model.

The LSTM consisted of two LSTM layers, a fully connected layer and an output layer. The first LSTM layer contained 10 LSTM cells, each with 128 neurons. The second LSTM layer also contained 10 LSTM cells, but each cell had 64 neurons. The fully connected layer had 54 neurons, and the final output layer had two neurons that output the valence and arousal scores. Each of the abovementioned layers used a dropout rate of 0.5, which adopted the ***ReLU*** activation function and required data normalization. The mean square error was used as the loss metric for this LSTM network.

Fusion Methods

By recording EEG signals and facial images through hardware devices (Emotiv EPOC+ headset and optical computer camera), MindLink-Eumpy toolbox reads and saves multimodal data and combines the predictions of the SVM (EEG modality) and CNN (facial expression modality) in decision level to improve the accuracy of emotion recognition. This subsection describes two decision-level fusion methods: the weight enumerator and AdaBoost method.

Weight Enumerator

We designed an enumerator to traverse weights in steps of 0.01 and find the optimal weights for the linear combination of two sub-classifiers. Equation (5) defines the linear combination.

$$S_{enum} = \sigma S_{face} + (1 - \sigma)\, S_{EEG} \tag{5}$$

In equation (5), σ ranges from 0 to 1, which represents the importance degree of the facial expression classifier; $S_{face}$and S_{EEG} represent the prediction scores of the facial expression classifier and EEG-based classifier, respectively. The value of σ that achieves the highest accuracy is selected as the optimal weight for linear combination. Equation (6) defines the combined emotion classification relations.

$$r_{enum} = \begin{cases} high & S_{enum} \geq 0.5 \\ low & S_{enum} < 0.5 \end{cases} \tag{6}$$

MindLink-Eumpy separately applies this fusion method in both the valence and arousal dimensions to classify emotion into four states.

AdaBoost

The second fusion method we used is the AdaBoost technique, which is to obtain the best parameters of $\omega_j (j=1,2,\ldots,n)$ for sub-classifiers. Equations (7) and (8) show the core mathematical formulas of AdaBoost.

$$S_{boost} = \frac{1}{\left(1 + \exp\left(-\sum_{j=1}^{n} w_j s_j\right)\right)} \tag{7}$$

$$r_{boost} = \begin{cases} high & S_{boost} \geq 0.5 \\ low & S_{boost} < 0.5 \end{cases} \tag{8}$$

In equation (7) and the below equations (9), (10), (11), (12), and (13), n represents the number of sub-classifiers, $s_j \in \{-1, 1\}_{(j=1,2,\ldots,n)}$ designates the outputs of the$_{jth}$ sub-classifier for the$_{ith}$ sample, and s_{boost} represents the scores of fused emotion regression, which are calculated by the AdaBoost algorithm. For example, in this study, $s_1$represents an EEG-based sub-classifier and $_{s_2}$ represents a facial expression sub-classifier. In equation (8), r_{boost}represents the fused emotion classification result.

The main process of AdaBoost is as follows. First, the training weights are initialized, as shown in equation (9):

$$\alpha_i = \frac{1}{m} \tag{9}$$

$\alpha_i$in equation (9) represents the weight of the $_{ith}$ sample, and $\boldsymbol{m}$ in equations (9) and (13) represents the size of the training sample. Each time AdaBoost updates sub-classifiers during model training, the sample data should be multiplied by the weights updated in the previous sub-classifier step. Equation (10) shows the mathematical formula for the error rate$_{\varepsilon_j}$.

$$\varepsilon_j = \sum_{i=1}^{M} t_i \alpha_i \tag{10}$$

In equation (10), $t_i$is calculated from equation (11), and $_{y_i}$ in equation (11) denotes the $_{ith}$ ground-truth label.

$$t_i = \begin{cases} 0 & s(x_i)_j = y_i \\ 1 & s(x_i)_j \neq y_i \end{cases} \tag{11}$$

Then, we calculate the weights of the sub-classifier using equation (12).

$$w_j = \frac{1}{2} \ln\left(\frac{1-\varepsilon_j}{\varepsilon_j}\right) \tag{12}$$

Next, we update weights for the next sub-classifier based on equation (13),

$$\alpha_{j+1,i} = \begin{cases} \frac{\alpha_{j,i} \exp(-w_j)}{\sum_{i=1}^{m} \alpha_{j,i} \exp(-w_j)} & s(x_i)_j = y_i \\ \frac{\alpha_{j,i} \exp(w_j)}{\sum_{i=1}^{m} \alpha_{j,i} \exp(w_j)} & s(x_i)_j \neq y_i \end{cases} \tag{13}$$

where $\boldsymbol{j}$ represents the $_{jth}$ sub-classifier, α represents the weight of the $_{ith}$ sample for the $_{jth}$ sub-classifier, and $_{s(x)_j}$ represents outputs of the $_{jth}$ sub-classifiers for the $_{ith}$ sample.

MindLink-Eumpy also separately applies this fusion method in both the valence and arousal dimensions to classify emotion into two states (low or high state in each dimension).

EXPERIMENTS AND RESULTS

This section describes the experiments performed to evaluate MindLink-Eumpy. Three experiments were conducted in this study, including two offline experiments and one online experiment. In the offline experiments, for each database, subjects were selected according to the falling criteria: **(i)** the subject's data contain both EEG and facial images; **(ii)** the subject's ground-truth labels contain two states, including low and high in both valence and arousal dimensions.

Offline Analysis

Experiments for the Subject-Dependent Methods

In this experiment, we used the DEAP database and the MAHNOB-HCI database to demonstrate the effectiveness of the subject-dependent methods based on multimodal emotion recognition. We chose 10 subjects in the DEAP database and 14 subjects in the MAHNOB-HCI database, then for each database, we randomly selected 20 trials of data of each subject as the training datasets, and the remaining 20 trials were used as the test datasets. **Figure 7** and **Table 2** show the offline experimental results (average values and accuracy thresholds) for the subject-dependent models in the DEAP database; and **Figure 8** and **Table 3** show the results in the MAHNOB-HCI database. The experimental results show that the first fusion method, the weight enumerator, achieved the highest accuracy in both the valence (in the DEAP database and the MAHNOB-HCI database) and arousal dimensions (in the MAHNOB-HCI database). The second fusion method, AdaBoost, also had a promising average accuracy, but the overall performance was lower than that of the weight enumerator, which was probably because of the less number of sub-classifiers (only has two modalities including facial expression and EEG signals). In addition, among the single-modal methods (that is, when there is only facial emotion recognition or EEG-based emotion recognition), the SVM in the valence dimension and CNN in the arousal dimension displayed promising accuracy. However, single-modal emotion recognition method was still less stable than the multimodal method. In this experiment, we observed that most subjects used the multimodal method to obtain higher accuracy, especially subjects 4 and 5 in the valence dimension in the MAHNOB-HCI and subjects 13 and 14 in the arousal dimension in the MAHNOB-HCI.

Furthermore, we conducted a normality test for these four methods (the SVM, CNN, weight enumerator, and AdaBoost methods). The data samples were considered normally distributed when the result was below 0.05; otherwise, we conducted another paired t-test procedure. During the t-test procedure, we considered that when the $\boldsymbol{p}$ value was lower than 0.05, the difference was statistically significant. The experiments were conducted based on the DEAP database and MAHNOB-HCI database. For the DEAP database, in the valence dimension, there was not only a significant difference ($\boldsymbol{p} < 0.01$) between the enumerator-based fusion results (weight enumerator) and the EEG-based results but also a significant difference ($\boldsymbol{p} = 0.016$) between the AdaBoost fusion results and the EEG-based results, but no significant difference was observed in the arousal dimension. For the MAHNOB-HCI database, no significant difference was observed in the valence dimension; but in the arousal dimension, there was a significant difference ($\boldsymbol{p} = 0.045$) between the AdaBoost fusion results and the EEG-based results.

Experiments for the Subject-Independent Methods

In the experiment for subject-independent methods, we used the EEG dataset collected from 30 subjects in the MAHNOB-HCI database to train and evaluate the LSTM model. We conducted the experiment by the following steps: Data from subject 1 to subject 20 were selected as the training set. After model training, we conducted an evaluation test using all data from subject

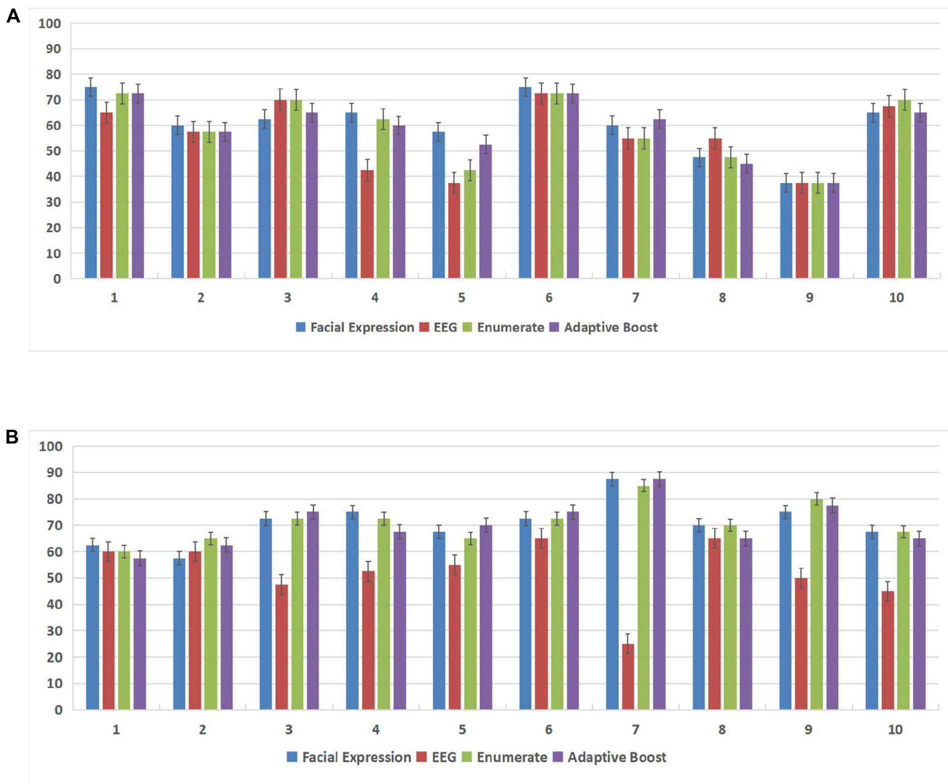


FIGURE 7 | Accuracy for each subject in both the arousal and valence dimensions in the Database for Emotion Analysis Using Physiological Signals (DEAP) database. The X-axis of each subfigure represents the subject ID, and the Y-axis represents the accuracy (%). **(A)** Accuracy (%) for each subject in the arousal dimension in the DEAP database. **(B)** Accuracy (%) for each subject in the valence dimension in the DEAP database.

TABLE 2 | Average accuracy (%) of subject-dependent models based on the DEAP database.

Target	Facial expression	EEG	Enumerator fusion	AdaBoost fusion
Valence	70.75 ± 7.67	52.50 ± 11.29	71.00 ± 7.00	70.25 ± 8.25
Arousal	60.50 ± 10.83	56.00 ± 12.46	58.75 ± 12.26	59.00 ± 10.74

DEAP, Database for Emotion Analysis Using Physiological Signals; EEG, electroencephalogram.

21 to subject 23. The experimental results show that the well-trained LSTM model achieved an accuracy of 78.56% and a recall rate of 68.18% in the valence dimension. Meanwhile, it achieved an accuracy of 77.22% and a recall rate of 69.28% in the arousal dimension. All the experimental results are shown in **Table 4**, including the training losses (Loss), validation loss, accuracy, recall rate, and root mean square error (RMSE). **Table 4** shows that values of training losses for the valence and arousal dimensions were 3.16 and 2.17, respectively, and the validation losses were 3.35 and 3.30, respectively. Furthermore, the recall rates for the valence and arousal dimensions were 68.18% and 69.28%, respectively, and the RMSEs were 1.83 and 1.82, respectively.

Online Experiment

In the online experiment, we used the Emotiv EPOC+ headset and optical computer camera to record EEG data and facial images. Fifteen healthy subjects participated in the experiment, including eight males and seven females. The ages of the subjects ranged from 17 to 21 years old (mean = 20.27, SD = 1.24). Before the experiment, 40 videos for emotion elicitation were selected from YouTube. We manually divided these videos into two groups for calibration and evaluation experiments. The video clips ranged in duration from 70.52 to 195.12 s (mean = 143.04, SD = 33.50). During the experiments, we calibrated the position of the headset and the camera and ensured that the subjects were in a comfortable environment. Then, subjects were instructed to watch emotion-eliciting video clips and stay focused, remain calm, and avoid blinking or moving during the viewing process. After the end of each experimental trail, the subjects reported their emotion status in the valence and arousal dimensions through a questionnaire.

In the calibration process, we conducted experiments for data collection and ground-truth label calibration. We performed 20 trials for each subject. For the convenience of extracting data from specific subjects, each subject was required to provide

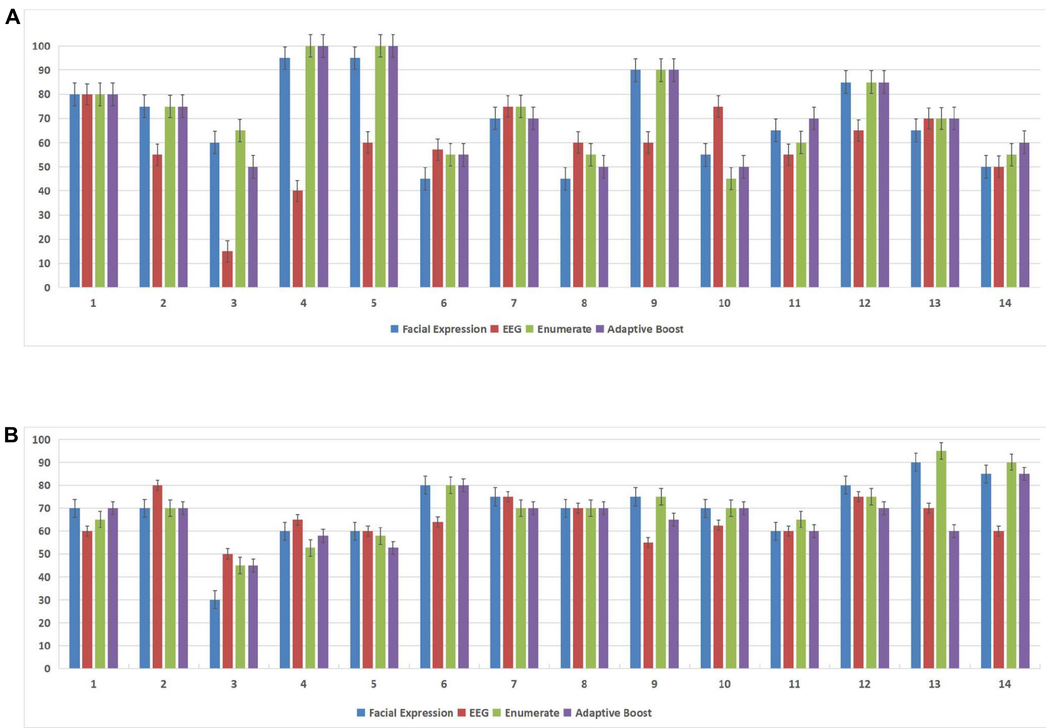


FIGURE 8 | Accuracy for each subject in both the arousal and valence dimensions in the MAHNOB-HCI database. The X-axis of each subfigure represents the subject ID, and the Y-axis represents the accuracy (%). **(A)** Accuracy (%) for each subject in the arousal dimension in the MAHNOB-HCI database. **(B)** Accuracy (%) for each subject in the valence dimension in the MAHNOB-HCI database.

TABLE 3 | Average accuracy (%) of subject-dependent models based on the MAHNOB-HCI database.

Target	Facial expression	EEG	Enumerator fusion	AdaBoost fusion
Valence	69.64 ± 14.07	64.75 ± 8.05	70.04 ± 12.81	66.11 ± 10.04
Arousal	69.64 ± 16.95	58.36 ± 15.86	72.14 ± 16.77	71.79 ± 16.97

EEG, electroencephalogram.

personal information before the start of the first experimental trial so that MindLink-Eumpy toolbox could associate the information of each subject with the corresponding physiological data and ground-truth labels, such as name, age, and gender. At the beginning of each trial, a 10-s countdown appeared in the center of the computer screen to attract the subject's attention. After the countdown, a video was presented in the screen to elicit the subject's emotion. MindLink-Eumpy recorded four facial images and 10 groups of EEG signals per second and then saved the data in the database. At the end of each trial, each subject was required to assign SAM scale values for the valence and arousal scores. After clicking the "submit" button, the next trial started, and a 10-s countdown appeared again between the adjacent trials. **Figure 9** presents the workflow of one trial for data collection in the online experiment.

In the evaluation experiments, a similar experimental trial process was used to evaluate the models. We used different videos to elicit the subject's emotion. In each trial, we used four methods (the SVM, CNN, weight enumerator, and AdaBoost methods) to detect emotion. We calculated accuracy by comparing the predicted emotions and ground-truth labels.

Figure 10 and **Table 5** show the online experimental results (average values and thresholds of accuracy) for models of subject dependence. Notably, the multimodal methods achieved higher accuracy than the single-modal methods, except that the accuracy of the EEG-based SVM in the arousal dimension was higher than that of the enumerated fusion method.

We also conducted a paired t-test for the online experiments. According to the experimental results, significant differences were observed both in valence and arousal dimensions between the AdaBoost fusion results and the facial expression results, where $\boldsymbol{p} = 0.014$ in the valence dimension and $\boldsymbol{p} = 0.049$ in the arousal dimension.

DISCUSSION AND CONCLUSION

Summary

This paper proposes MindLink-Eumpy, which is an open-source Python toolbox for multimodal emotion recognition. MindLink-Eumpy includes a series of tools for data collection, multimodal data processing, machine learning methods, and deep learning

TABLE 4 | Experimental results of the subject-independent model on MAHNOB-HCI.

Dimension	Loss	Validation loss	Accuracy	Recall rate	RMSE
Valence	3.16	3.35	78.56%	68.18%	1.83
Arousal	2.17	3.30	77.22%	69.28%	1.82

RMSE, Root mean square error.

methods. Our aim of developing MindLink-Eumpy is to provide an extensible software framework for research and application in the field of emotion recognition.

First, MindLink-Eumpy implements an event-related paradigm that uses videos to elicit subject's emotion. To record facial images and EEG information for subjects, MindLink-Eumpy implements tools to invoke computer cameras and headsets. Most importantly, MindLink-Eumpy implements a series of methods to overcome traditional difficulties in the field of emotion recognition: (i) two fusion methods are applied to combine SVM (EEG modality) and CNN (facial images) results to improve the accuracy of emotion recognition; (ii) a multitask CNN (facial images) based on transfer learning is used to overcome the overfitting phenomenon caused by lacking of image data; and (iii) LSTM based on EEG is used to implement a subject-independent emotion recognition technique. Finally, all the above methods have been tested and demonstrated to be effective. In particular, in experiments for the subject-independent methods, although the performance deteriorates when training data and validation data were completely heterogeneous, an acceptable accuracy was still maintained. Thus, this method has promising stability. Although MindLink-Eumpy is still in its infancy, it has the potential to become a benchmark toolbox in industrial and lab applications in the emotion recognition area.

Analysis of the Advantages of MindLink-Eumpy

Advantages of MindLink-Eumpy mainly consist of three points: **(i)** a promising accuracy and subject independence for emotion recognition, **(ii)** a better robustness and scalability for software, **(iii)** a framework for online experimental paradigm and data storage medium.

Specifically, in this study, we proposed two approaches including facial expression and EEG for multimodal emotion recognition, each of which corresponds to a sub-classifier. Facial expression recognition approach is more accurate, but users may camouflage expressions in the real usage scenario, so an EEG emotion recognition approach is in use to fill the gap between the error associated with facial expressions and the Bayesian error in ground-truth emotion labels. To fuse multimodal information, we proposed two methods, weight enumerator and Adaboost, to improve the emotion recognition accuracy. Furthermore, we proposed another approach for subject-independent emotion recognition (based on LSTM in EEG modality) to make it suitable for more users. The subject-independent approach is independent of SVM and CNN approaches. Meanwhile, although feature-level fusion is more accurate theoretically, it is difficult to fuse spatial features (such as facial images) with temporal features (such as EEG information). Therefore, we applied decision-level fusion methods to ensure a better robustness mentioned above, which means that MindLink-Eumpy is able to keep running steadily even if there occurs errors in one equipment (such as camera, brain–computer interfaces, or other devices to be added). It is easier to add different modality information for multimodal emotion recognition using the methods of decision-level fusion in MindLink-Eumpy. Furthermore, MindLink-Eumpy provides a framework for online experiments. During the online experiments, MindLink-Eumpy stores physiological data into a folder system for future scientific research.

Although the performance was improved by information fusion, the superiority of the multimodal fusion over the single-modal approach did not show strong statistical significance in our results (e.g., when an independent two-sample t-test was performed on the accuracy distribution, the p values in **Tables 2**, **3** are not always less than 0.05). In many emotion experiments [e.g., Alsolamy and Fattouh (2016)], it can be found that high volatility is associated with facial expressions because subjects can trick the machine by imitating certain facial expressions. For this problem, the gap between the error related to facial expressions and the Bayesian errors of true emotion detection generally can be filled by adding information sources (e.g., EEG) (Li et al., 2019). For the experiments on DEAP and MAHNOB-HCI databases, the subjects were asked to behave normally rather than mimic certain facial expressions, which may be the main reason that we could not find strong statistical evidence indicating significant improvement after fusion. Furthermore, there are only 10 subjects in the DEAP and 14 subjects in the MAHNOB-HCI that meet our experimental requirements. The limited sample size may be another reason why the results are not statistically significant.

Comparison With Other Methods in the Literature

The main functions of MindLink-Eumpy comprise data collection and storage, data preprocessing, feature extraction and visualization, and emotion recognition. This toolbox is integrated with our methods for multimodal emotion recognition. Moreover, it is able to facilitate scientific research on multimodal emotion recognition. Compared with other studies, MindLink-Eumpy complements the existing research in some ways.

From the perspective of multimodal emotion recognition methods, it is important to achieve a promising accuracy and make models subject independent. MindLink-Eumpy combines facial expression and EEG for a promising accuracy and provides an LSTM model based on EEG for subject independence. Li et al. (2019) combined EEG and facial expression data to optimize emotion recognition algorithms. Three types of movie

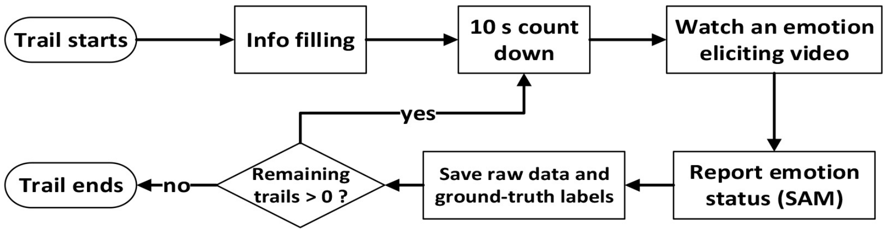


FIGURE 9 | Workflow for one subject in the online experiment.

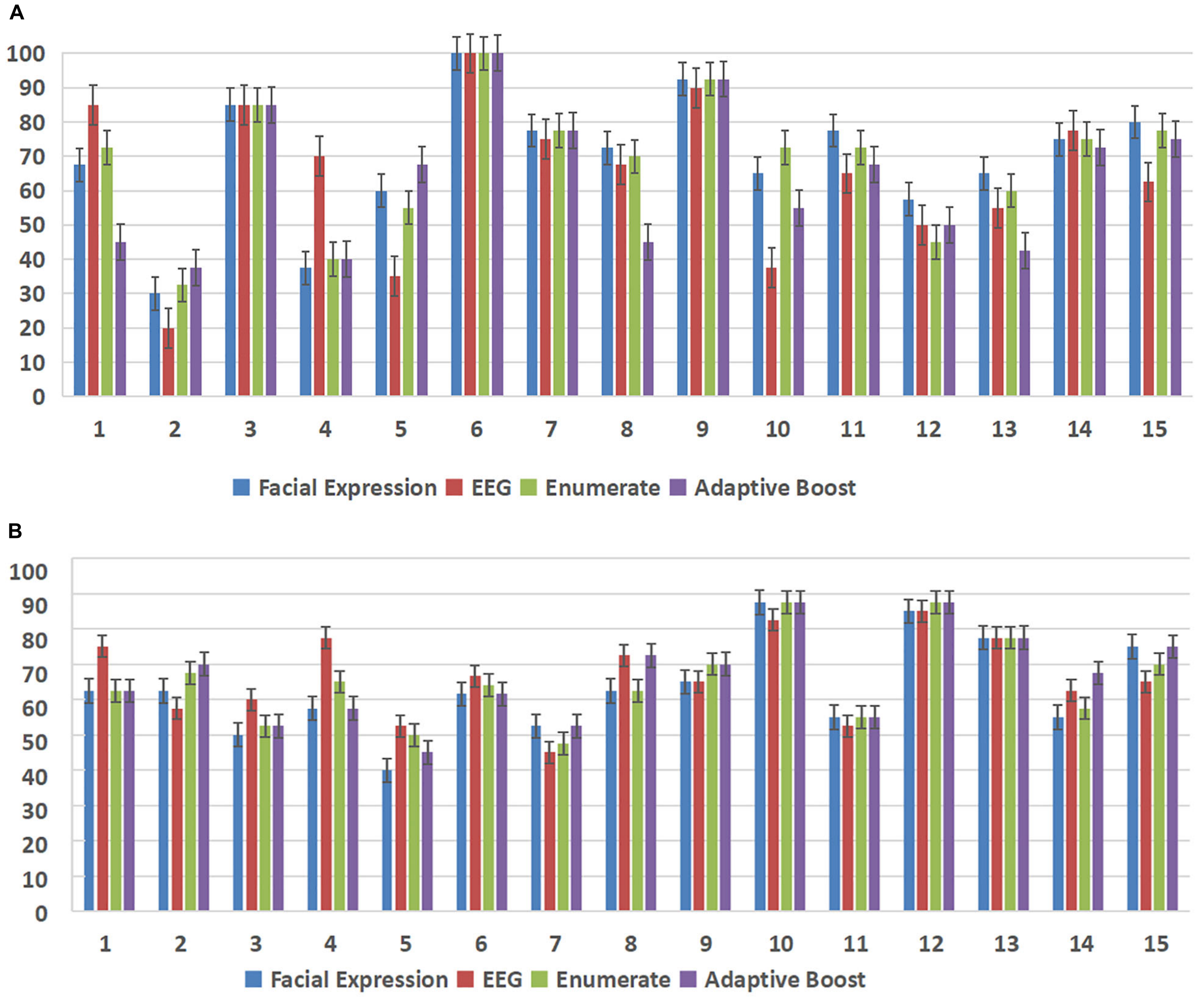


FIGURE 10 | Accuracy (%) of each subject in both the arousal and valence dimensions in the online experiment. The X-axis of each subfigure represents the subject ID, and the Y-axis represents the accuracy (%). **(A)** Accuracy (%) of each subject in the arousal dimension in the online experiment. **(B)** Accuracy (%) of each subject in the valence dimension in the online experiment.

clips (positive, neutral, and negative) were utilized for emotion data collection, and LSTM was utilized for decision-level fusion and capturing temporal dynamics of emotion, which yield a concordance correlation coefficient (CCC) of 0.625 ± 0.029.

Under the circumstance that we divide emotions into two classifications in both valence and arousal dimension (four categories in total), MindLink-Eumpy is demonstrated to be more accurate and subject independent in emotion recognition.

TABLE 5 | Average (%) accuracy of the subject-dependent models for online databases.

Target	Facial expression	EEG	Enumerate fusion	AdaBoost fusion
Valence	63.27 ± 12.68	66.44 ± 11.39	65.12 ± 11.78	66.27 ± 12.23
Arousal	69.50 ± 17.96	65.00 ± 21.53	68.50 ± 18.28	63.50 ± 19.47

EEG, electroencephalogram.

In a binary classification of valence and arousal dimensions, our experimental results demonstrated that MindLink-Eumpy has a promising performance in both offline analysis and real-time detection. For instance of experiments on MAHNOB-HCI database, the average accuracy of 68.50% was achieved in arousal dimension in the study of Koelstra and Patras (2013), while in our study, it reached 72.14%. Taking the DEAP database as an example, the study of He et al. (2017) achieved an average accuracy of 70.90% in valence dimension, while the average of accuracy in our study reached 72.14%. Chen et al. (2017) combined EEG information, peripheral physiological signals, and facial video to obtain promising accuracy (77.57% for four emotion classifications). This study obtained higher accuracy and subject independence than the work of Buitelaar et al. (1999). However, it lacked facial expression modality and needed wearable equipment to acquire all of the physiological signals, which means that it is stringent to apply their method in practice.

Furthermore, from the perspective of data and software, not only research on emotion recognition needs more physiological data storage medium, feature extractors, and experimental framework, but practical application scenario needs functions of data streaming and real-time visualization. MindLink-Eumpy provides a folder system for data storage and integrates a series of tools for preprocessing and feature extraction of facial images and EEG signals. Besides, MindLink-Eumpy is suitable for devices including brain–computer interfaces and cameras for real-time data acquisition and visualization. Buitelaar et al. (2018) proposed a toolbox named MixedEmotion that provided audio processing, text processing, and video processing for multimodal emotion analysis. MixedEmotion is well developed and practical application oriented, but it lacks physiological information for intuitive feedback. It mainly focuses on enterprise application but does not contribute to scientific research on human neuroscience. Soleymani et al. (2017) proposed a toolbox named TEAP for the signal processing of EEG, GSR, EMG, skin temperature, respiration pattern, and blood volume pulse information, which expanded the application scope of multimodal emotion analysis. The authors of Soleymani et al. (2017) had tried to replicate some methods of other articles and demonstrate the effectiveness of feature extraction function of TEAP. But they had not proposed their original methods for emotion recognition. Generally speaking, MindLink-Eumpy provides a framework for scientific research and application with our original approaches for subject-independent emotion recognition. Compared with other works, MindLink-Eumpy promotes research in the area of emotion recognition.

Compared with a subject-dependent approach, LSTM-based achieved higher accuracies. Here are two reasons for this situation. First, the inherent difference between SVM and LSTM may be one of the reasons for the performance difference between them. A previous study (Nath et al., 2020) compared the different performances of LSTM in subject-dependent and subject-independent experiments. Their results in subject-independent recognition showed that the average accuracy rates of valence and arousal dimensions were 70.31% and 69.53%, respectively. The experimental results are basically consistent with our conclusion that the LSTM-based method achieved an accuracy of 78.56% in the valence dimension and an accuracy of 77.22% in the arousal dimension for the subject-independent recognition. Second, the outliers of data also hinder the high performance of subject-dependent methods. In this study, although we paid much attention to data preprocessing, we still found some strange but not dirty data different from normal ones. For example, the PSD values of some trails of certain subjects remain low. It is hard to filter all these outliers of data for SVM. Therefore, in the case of the subject dependence, a well-trained SVM model may pay too much attention to outliers, thereby reducing the imitation effect of the model. For the case of the subject independence, abundant data enable the LSTM model to focus on the universality of features of all subjects rather than outliers, thereby eliminating the effects caused by relatively few outliers. In the future, we will try to analyze the abnormal data and remove outliers.

Potential Applications

MindLink-Eumpy could provide a potential software benchmark for emotion recognition in industry applications. Thus, there are various potential applications based on the technologies and frameworks of MindLink-Eumpy. In the medical field, emotion recognition plays an important role in the treatment of children with autism (Buitelaar et al., 1999), hearing-impaired children (Gu et al., 2019), patients with depression (Punkanen et al., 2011), etc., In the field of intelligent driving, research has focused on the behaviors of drivers affected by emotions (Roidl et al., 2014). Extreme emotions might lead to improper operations or even traffic accidents, thereby endangering drivers and passengers. Emotion recognition technologies can also be applied for supervisory care, including baby care, intensive care, Alzheimer's care, etc., Overall, MindLink-Eumpy has promising application prospects.

Limitations

As a software toolbox, MindLink-Eumpy has limitations. First, MindLink-Eumpy only provides tools for EEG information and facial images. Other widely used physiological data, such as eye movement signals and electrocardiograph (ECG) information, are not currently compatible with MindLink-Eumpy. In addition, although multimodal emotion recognition methods outperform single-modal methods and the average accuracy of multimodal methods is high, the results of our experiments did not display strong statistical significance. In the experiments, as described by the Hawthorne effect, subjects may tend to display biased facial expressions due to their awareness of being observed. Moreover, it is challenging to record a large EEG dataset because of the volatility of conductive media in brain-computer interface (BCI) and the lack of subjects.

Future Work

In the future, we will attempt to add new data modalities to improve multimodal emotion recognition and implement new tools suitable for different hardware devices. In single-modal EEG-based emotion recognition, we will implement semi-supervised machine learning algorithms for cross-subject detection. Furthermore, we will try to use eye movement signal to measure domain differences among subjects and implement methods for the feature-level fusion of eye movement signals and EEG information.

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 Ethics Committee of South China Normal University. 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

JP: conceptualization, supervision, project administration, and funding acquisition. JP, RL, and YL: methodology and validation. RL, ZC, and BW: software. RL, WH, and ZC: formal analysis. JP, LQ, and YL: investigation. RL and XL: writing – original draft preparation. JP, YL, YY, and LQ: writing – review and editing. RL, WH, and BW: visualization. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Key Realm R&D Program of Guangzhou under grant 202007030005, the National Natural Science Foundation of China under grant 62076103, the Guangdong Natural Science Foundation under grant 2019A1515011375, the Special Funds for the Cultivation of Guangdong College Students' Scientific and Technological Innovation under grant pdjh2020a0145, and the Guangdong Basic and Applied Basic Research Foundation under grant 2019A1515110388.

ACKNOWLEDGMENTS

We would like to thank Yongrui Huang, Jianhao Yang, and Siyu Liu for technical assistance. We gratefully acknowledge the developers of Python, Tensorflow, Keras, scikit-learn, NumPy, Pandas, MNE, and other software packages that MindLink-Eumpy builds upon.

REFERENCES

Alhagry, S., Aly, A., and El-Khoribi, R. (2017). Emotion recognition based on EEG using LSTM recurrent neural network. *Int. J. Adv. Comput. Sci. Appl.* 8, 355–358. doi: 10.14569/IJACSA.2017.081046

Alsolamy, M., and Fattouh, A. (2016). "Emotion estimation from EEG signals during listening to Quran using PSD features," in *Proceedings of the 2016 7th International Conference on Computer Science and Information Technology (CSIT)* (Amman: IEEE), 1–5. doi: 10.1109/CSIT.2016.7549457

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

Buitelaar, J., Wees, M., Swaab, H., and Gaag, R. (1999). Theory of mind and emotion-recognition functioning in autistic spectrum disorders and in psychiatric control and normal children. *Dev. Psychopathol.* 11, 39–58. doi: 10.1017/S0954579499001947

Buitelaar, P., Wood, I. D., Negi, S., Arcan, M., McCrae, J. P., Abele, A., et al. (2018). Mixed emotions: an open-source toolbox for multimodal emotion analysis. *IEEE Trans. Multimed.* 20, 2454–2465. doi: 10.1109/TMM.2018.2798287

Castellano, G., Kessous, L., and Caridakis, G. (2008). "Emotion recognition through multiple modalities: face, body gesture, speech," in *Affect and Emotion in Human-Computer Interaction: From Theory to Applications Lecture Notes in Computer Science*, eds C. Peter and R. Beale (Berlin: Springer), 92–103. doi: 10.1007/978-3-540-85099-1_8

Chen, J., Hu, B., Wang, Y., Moore, P., Dai, Y., Feng, L., et al. (2017). Subject-independent emotion recognition based on physiological signals: a three-stage decision method. *BMC Med. Inform. Decis. Mak.* 17:167. doi: 10.1186/s12911-017-0562-x

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297. doi: 10.1007/BF00994018

Damasio, A. R. (1998). Emotion in the perspective of an integrated nervous. *Brain Res. Rev.* 26, 83–86. doi: 10.1016/S0165-0173(97)00064-7

Das, P., Sadhu, A. K., Konar, A., Bhattacharya, B. S., and Nagar, A. K. (2015). "Adaptive parameterized AdaBoost algorithm with application in EEG motor imagery classification," in *Proceedings of the 2015 International Joint Conference on Neural Networks (IJCNN)*, (Killarney: IEEE), 1–8. doi: 10.1109/IJCNN.2015.7280835

Davidson, R. J., Ekman, P., Saron, C. D., Senulis, J. A., and Friesen, W. V. (1990). Approach-withdrawal and cerebral asymmetry: emotional expression and brain physiology: I. *J. Pers. Soc. Psychol.* 58, 330–341. doi: 10.1037/0022-3514.58.2.330

Di Flumeri, G., Arico, P., Borghini, G., Sciaraffa, N., Maglione, A. G., Rossi, D., et al. (2017). "EEG-based approach-withdrawal index for the pleasantness evaluation during taste experience in realistic settings," in *Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (Seogwipo: IEEE), 3228–3231. doi: 10.1109/EMBC.2017.8037544

Duan, K.-B., Rajapakse, J. C., Wang, H., and Azuaje, F. (2005). Multiple SVM-RFE for gene selection in cancer classification with expression data. *IEEE Trans. Nanobiosci.* 4, 228–234. doi: 10.1109/tnb.2005.853657

Georgieva, O., Milanov, S., Georgieva, P., Santos, I. M., Pereira, A. T., and Silva, C. F. (2015). Learning to decode human emotions from event-related potentials. *Neural Comput Appl.* 26, 573–580. doi: 10.1007/s00521-014-1653-6

Goodfellow, I. J., Erhan, D., Luc Carrier, P., Courville, A., Mirza, M., Hamner, B., et al. (2015). Challenges in representation learning: a report on three machine learning contests. *Neural Netw.* 64, 59–63. doi: 10.1016/j.neunet.2014.09.005

Gu, H., Fan, R., Zhao, J., Chen, Y., Chen, Q., and Li, X. (2019). Inhibitory control of emotional interference in children with learning disorders: evidence from event-related potentials and event-related spectral perturbation analysis. *Brain Res.* 1718, 252–258. doi: 10.1016/j.brainres.2019.04.016 s

Guzel Aydin, S., Kaya, T., and Guler, H. (2016). Wavelet-based study of valence–arousal model of emotions on EEG signals with LabVIEW. *Brain Inf.* 3, 109–117. doi: 10.1007/s40708-016-0031-9

He, Y., Ai, Q., and Chen, K. (2017). "A MEMD method of human emotion recognition based on valence-arousal model," in *2017 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, Vol. 2, (Hangzhou: IEEE), 399–402. doi: 10.1109/IHMSC.2017.201

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780. doi: 10.1162/neco.1997.9.8.1735

Jiang, W., Liu, G., Zhao, X., and Yang, F. (2019). "Cross-subject emotion recognition with a decision tree classifier based on sequential backward selection," in *Proceedings of the 2019 11th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, (Hangzhou), 309–313. doi: 10.1109/IHMSC.2019.00078

Koelstra, S., Muhl, C., Soleymani, M., Lee, J., Yazdani, A., Ebrahimi, T., et al. (2012). DEAP: a database for emotion analysis; using physiological signals. *IEEE Trans. Affect. Comput.* 3, 18–31. doi: 10.1109/T-AFFC.2011.15

Koelstra, S., and Patras, I. (2013). Fusion of facial expressions and EEG for implicit affective tagging. *Image Vision Comput.* 31, 164–174. doi: 10.1016/j.imavis.2012.10.002

Lang, P., Bradley, M., and Cuthbert, B. (1997). "Motivated attention: affect, activation, and action," in *Attention and Orienting: Sensory and Motivational Processes*, eds P. J. Lang, R. F. Simons, and M. T. Balaban (New Jersey, NJ: Lawrence Erlbaum Associates Publishers), 97–135.

Lawrence, S., Giles, C. L., Tsoi, A. C., and Back, A. D. (1997). Face recognition: a convolutional neural-network approach. *IEEE Trans. Neural Netw.* 8, 98–113. doi: 10.1109/72.554195

Li, D., Wang, Z., Wang, C., Liu, S., Chi, W., Dong, E., et al. (2019). The fusion of electroencephalography and facial expression for continuous emotion recognition. *IEEE Access* 7, 155724-155736. doi: 10.1109/ACCESS.2019.2949707

Li, J., Qiu, S., Shen, Y.-Y., Liu, C.-L., and He, H. (2020). Multisource transfer learning for cross-subject eeg emotion recognition. *IEEE Trans. Cybern.* 50, 3281–3293. doi: 10.1109/TCYB.2019.2904052

Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., et al. (2011). "The computer expression recognition toolbox (CERT)," in *Proceedings of the 2011 IEEE International Conference on Automatic Face Gesture Recognition (FG)*, (Santa Barbara, CA), 298–305. doi: 10.1109/FG.2011.5771414

Nath, D., Anubhav, A., Singh, M., Sethia, D., Kalra, D., and Sreedevi, I. (2020). "A comparative study of subject-dependent and subject-independent strategies for EEG-based emotion recognition using LSTM network," in *Proceedings of the 2020 the 4th International Conference on Compute and Data Analysis*, (Silicon Valley CA), doi: 10.1145/3388142.3388167

Ng, W., Saidatul, A., Chong, Y. F., and Ibrahim, Z. (2019). PSD-based features extraction for EEG signal during typing task. *IOP Conf. Ser. Mater. Sci. Eng.* 557:012032. doi: 10.1088/1757-899X/557/1/012032

Nguyen, D., Nguyen, K., Sridharan, S., Abbasnejad, I., Dean, D., and Fookes, C. (2018). "Meta transfer learning for facial emotion recognition," in *Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR)*, (Beijing: IEEE), 3543–3548. doi: 10.1109/ICPR.2018.8545411

Park, S.-A., Hwang, H.-J., Lim, J.-H., Choi, J.-H., Jung, H.-K., and Im, C.-H. (2013). Evaluation of feature extraction methods for EEG-based brain–computer interfaces in terms of robustness to slight changes in electrode locations. *Med. Biol. Eng. Comput.* 51, 571–579. doi: 10.1007/s11517-012-1026-1

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830.

Poria, S., Cambria, E., Bajpai, R., and Hussain, A. (2017). A review of affective computing: from unimodal analysis to multimodal fusion. *Inf. Fusion* 37, 98–125. doi: 10.1016/j.inffus.2017.02.003

Posner, J., Russell, J., and Peterson, B. (2005). The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev. Psychopathol.* 17, 715–734. doi: 10.1017/S0954579405050340

Prakash, R. M., Thenmoezhi, N., and Gayathri, M. (2019). "Face Recognition with convolutional neural network and transfer learning," in *Proceedings of the 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, (Tirunelveli), 861–864. doi: 10.1109/ICSSIT46314.2019.8987899

Punkanen, M., Eerola, T., and Erkkilä, J. (2011). Biased emotional recognition in depression: perception of emotions in music by depressed patients. *J. Affect. Disord.* 130, 118–126. doi: 10.1016/j.jad.2010.10.034

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

Roidl, E., Frehse, B., and Höger, R. (2014). Emotional states of drivers and the impact on speed, acceleration and traffic violations-a simulator study. *Accid. Anal. Prev.* 70, 282–292. doi: 10.1016/j.aap.2014.04.010

Sebe, N., Cohen, I., Gevers, T., and Huang, T. S. (2005). "Multimodal approaches for emotion recognition: a survey," in *Proceedings of the SPIE*, eds S. Santini, R. Schettini, and T. Gevers (San Jose, CA: Society of Photo-Optical Instrumentation Engineers), 56–67. doi: 10.1117/12.600746

Soleymani, M., Lichtenauer, J., Pun, T., and Pantic, M. (2012). A multimodal database for affect recognition and implicit tagging. *IEEE Trans. Affect. Comput.* 3, 42–55. doi: 10.1109/T-AFFC.2011.25

Soleymani, M., Villaro-Dixon, F., Pun, T., and Chanel, G. (2017). Toolbox foremotional feature extraction from physiological signals (TEAP). *Front. ICT* 4, 1–7. doi: 10.3389/fict.2017.00001

Sur, S., and Sinha, V. K. (2009). Event-related potential: an overview. *Ind. Psychiatry J.* 18, 70–73. doi: 10.4103/0972-6748.57865

Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *Proc. IEEE* 70, 1055–1096. doi: 10.1109/PROC.1982.12433

van der Walt, S., Colbert, S. C., and Varoquaux, G. (2011). The numpy array: a structure for efficient numerical computation. *Comput. Sci. Eng.* 13, 22–30. doi: 10.1109/MCSE.2011.37

Viola, P., and Jones, M. (2001). "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001* (Kauai, HI: IEEE Computer Society), I–511–I–518. doi: 10.1109/CVPR.2001.990517

Yan den Broek, E. L. (2013). Ubiquitous emotion-aware computing. *Pers. Ubiquit. Comput.* 17, 53–67. doi: 10.1007/s00779-011-0479-9

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

frontiers in Computer Science

OPINION
published: 24 March 2021
doi: 10.3389/fcomp.2021.628773

The Use of Brain–Computer Interfaces in Games Is Not Ready for the General Public

*Grégoire Cattan**

IBM, Krakow, Poland

Keywords: brain–computer interfaces, games, virtual reality, electroencephalography, perspectives

OPEN ACCESS

Edited by:
Anton Nijholt,
University of Twente, Netherlands

Reviewed by:
Leonardo Cunha de Miranda,
Federal University of Rio Grande do Norte, Brazil
Gabriel Vasiljevic,
Federal University of Rio Grande do Norte Natal, Brazil, in collaboration with reviewer LM
Marvin Andujar,
University of South Florida, United States

****Correspondence:***
Grégoire Cattan
gregoire.cattan@ibm.com

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Computer Science

Received: *12 November 2020*
Accepted: *02 March 2021*
Published: *24 March 2021*

Citation:
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

INTRODUCTION

The use of brain–computer interfaces (BCIs) based on electroencephalography (EEG) in video games has been widely investigated. Research in adaptive training, single-trial classification (Congedo, 2013; Barachant and Congedo, 2014) and the creation of affordable EEG acquisition devices (Vos et al., 2014; Yohanandan et al., 2018) has paved the way for the development of a ubiquitous BCI technology. For example, Congedo (2013) developed "Brain Invaders," a BCI game inspired by the famous vintage game *Space Invaders* (Taito, Tokyo, Japan) and based on the so-called visual P300—an electromagnetic potential produced by the brain about 300 ms after a visual stimulation. Brain Invaders uses an adaptive algorithm that allows the player to plug the material and play without the need for calibration (Barachant and Congedo, 2014), while still achieving a high accuracy rate (Barachant et al., 2012). The game also demonstrates a good understanding of game design by naturally incorporating visual stimulations in the virtual environment. In this respect, Kaplan et al. (2013), Cattan et al. (2018b), and Rashid et al. (2020) provided a set of guidelines to adapt game implementation for BCI games, such as the use of turn-based games with a slow gameplay. Although Brain Invaders use a research-grade amplifier, the feasibility of using a low-cost EEG acquisition system for BCI has been demonstrated by Vos et al. (2014) and Yohanandan et al. (2018). These affordable headsets are comparable with research-grade amplifiers. In addition, Lotte et al. (2008) and Debener et al. (2012) demonstrated promising results when BCIs were used out of the lab for BCIs based on visual stimulation and movement imagination. The feasibility of using BCIs outside the lab has also been demonstrated in different context and at events. For instance, in the BCI game developed by Mentalista (Paris, France) for the 2016 European Football Championship, two players were asked to score against each other by moving a ball toward the opposite player's cage by concentrating[1].

Although a positive step forwards, these achievements have led to the false opinion that BCIs are ready for entertainment – a belief that is supported by enthusiastic visions claiming, for example, that brains will be connected to the internet through USB[2]. This opinion is rather qualified in the scientific community, which reported that BCIs suffer from (1) a low transfer rate, (2) a lack of market-ready, affordable, and user-friendly research-grade EEG acquisition devices, and (3) a gap between the game design and graphics of video games available on the market vs. in laboratories. Priorities of these limitations for video game development are discussed in the literature (Nijholt et al., 2009; Ferreira et al., 2013; Marshall et al., 2013; van de Laar et al., 2013; Ahn et al., 2014; Cattan et al., 2018b; Kerous et al., 2018; Vasiljevic and Miranda, 2019; Pierce et al., 2020), and in general preponderance of quantitative *over* qualitative aspects is criticized (Nijholt et al., 2009; Vasiljevic and Miranda, 2019). This article supports the claim that BCIs are not ready for general public use, based on other aspects than performance. Limitations are further detailed in section Limitations of BCI Games of the present study, and the obstacles to public use are analyzed. The conclusion is presented in section Discussion and Conclusion.

[1]https://mentalista.fr/foot
[2]https://www.thequint.com/explainers/what-is-neuralink-and-how-does-it-work-explained

LIMITATIONS OF BCI GAMES

Low Transfer Rate

The transfer rate (in bits per second) is a computed measure for communication devices that reflects the speed and accuracy of a device. This measure is derived from information theory and was adapted for BCIs (Wolpaw and Wolpaw, 2012) to compare different implementations. Following the development of Riemannian Geometry during the last decade (e.g., Barachant et al., 2012), the transfer rate of BCIs has considerably increased to reach, for example in the case of a P300-based BCI, a 90% accuracy within a couple of seconds (Cattan et al., 2018b). For comparison, the Guinness World Records reports a record of stenotype writing of around 360 words per minutes with 97% accuracy[3]. In practice, this means that BCIs are unusable in traditional inputs, such as in keyboards or mice. Another concern is that the algorithm complexity [in terms of mathematics and set-up (e.g., Cattan et al., 2018a; Andreev et al., 2019)] behind BCIs might be an obstacle for game developers. This is true for VR games especially, as the VR market is mostly represented by small, independent companies with limited resources[4,5].

Cost and Encumberment of Materials

Until the early 2000s, research-grade amplifiers were medical and were made of expensive materials and could mostly only be afforded by institutions or consortia (e.g., g.USBAmp, g.tec, Schiedlberg, Austria). Emotiv (San Francisco, US) was among the first manufacturers to release a commercial EEG cap for individual customers in 2009. However, low-cost versions of the Emotiv headsets omit a proper electrode location for P300-based BCIs, while research-grade versions still practice dissuasive pricing. Similar concerns apply to most customer-grade EEGs (Ahn et al., 2014; Vasiljevic and Miranda, 2019). That is, cheap materials only include a few and unmovable channels which are inappropriate for BCI based on visual stimulation, whereas medical-grade EEG caps are expensive. In 2013, OpenBCI (New York, US) finished a successful fundraising campaign, with the aim of providing a high-quality EEG acquisition system for <1,000 euros. Nevertheless, concerns were raised about the usability of the technology, such as its association with unstable wireless communication, non-standardized sampling rate and use of gel electrodes (e.g., Chabance et al., 2019). Some researchers (e.g., Yohanandan et al., 2018) demonstrated that similar performance could be obtained with an in-house, and thus cheaper, EEG headset. A Huffington Post (New York, US) publication presented an EEG headset that was ergonomic (i.e., with dry electrodes and wireless) for <500 euros[6,7]. However, such headsets are not available for public use, in the sense that, even though some are open-source, developers cannot build them in practice.

Another consideration is that any development is material-dependant, as all are based on different hardware, drivers and protocols (e.g., Ahn et al., 2014; Pierce et al., 2020). Despite the lack of standards, commercial brands such as Emotiv (San Francisco, the US) or NeuroSky (San Rose, the US) have developed plugins for game development. For instance, Rosca and Leba (2019) developed a pool game by integrating the Emotiv SDK with Unity 3D (San Francisco, the US), a notorious game engine. A review of existing BCI software can be found in Pierce et al. (2020). That said, the lack of middleware supported by a large and independent community creates maintenance and portability issues, as long as plugins are constructor-dependent or rely on the willingness of a research team.

Lack of Game Design and Graphics

Graphics and game design are a key concern and expectation for games on the market, from AAA to indie games. This is why video game studios hire developers as well as graphic designers, art work designers, concept artists and game designers. However, graphics and game design have been underestimated in the development of BCI games in laboratories. For instance, the number of frames per second (FPS) is a major concern for stimulation-based BCI, as they require the exact onset of the stimulation with a precision of around ±2 ms (Andreev et al., 2016). Indeed, if the tagging of such stimulation on the ongoing EEG does not happen all the time in the same frame in which the stimulation is displayed, a jitter is observable, which varies as an inverse function of the frame rate. For example, an FPS of around 50 Hz will output a frame every 20 ms and generate a jitter of a similar amplitude in the worst cases. This is particularly true for VR games where low FPS can result in a higher jitter (Cattan et al., 2018a). However, to our knowledge, the impact of graphical quality in FPS resulting in jitters is poorly understood in the context of BCI entertainment.

Ahn et al. (2014) conducted an opinion survey on the importance of BCI games elements. One aspect of this study was to outline the difference of perspective between developers and researcher communities regarding BCI development. For example, easiness of playing was one of the most important elements for 58% of the developers but only 19% of the researchers. However, the importance of graphics was minimized in the study when it should be a major concern for *video* games. In fact, the authors reported a developer opinion (later confirmed by the opinion survey for the two communities) that aesthetic is rather not considered as one of the most important factors for BCI games. However, we believe this only means that graphics are a basic requirement for video games. In Schell (2014), a pillar reference for game design, the author said about one of his work experiences in virtual reality:

> *"We had to make things look beautiful. [. . .]. We used high-end graphics hardware and rich textures and models [. . .]."*

In general, laboratory BCI games demonstrate good design, being turn-based and having slow gameplay—in this aspect following

[3] https://www.guinnessworldrecords.com/world-records/fastest-realtime-court-reporter-(stenotype-writing)/
[4] https://www.vrfocus.com/2020/05/why-now-is-the-time-for-aaa-studios-to-consider-vr/
[5] https://labusinessjournal.com/news/2015/jun/17/independent-virtual-reality-studios-benefit-early-/
[6] https://www.youtube.com/watch?v=GgKEOlcX9R8
[7] http://alexandre.barachant.org/eeg.io/

the guidelines stated by the scientific community (e.g., Nijholt et al., 2009; Marshall et al., 2013). For example, Lotte et al. (2008) and Andreev et al. (2016) presented the BCI game *Use the force!*, which consisted of lifting up a vessel with motor imagery. Meanwhile, in the game *Alphawow* (van de Laar et al., 2013), the avatar's character changes its behavior according to the player's state of relaxation. These two games consider the low-transfer rate of the BCI by mapping it to a feature that is expected to fail from time to time (it can be agreed that using the mind to move objects is difficult to realize) and to not compete with traditional inputs (a keyboard cannot determine a person's relaxation state). Nevertheless, the use of BCI was restricted to a unique aspect of these games. On one hand, if this aspect is a side aspect (e.g., Alphawow), the use of BCI is not valuable because of its cost. On the other hand, if this aspect is the main aspect of the game, it means that the player's ability to finish the game depends on an unreliable input and thus leads to frustration. A subjective study on the use of BCI for gaming (Cattan et al., 2019) also reported a lack of feedback for error quantification from participants. This couples with the fact that around 20% of the users are not proficient using a typical BCI. In fact, BCI illiteracy is an issue which is also well-established in the existing literature (e.g., Nijholt et al., 2009; Allison and Neuper, 2010; Marshall et al., 2013; Vasiljevic and Miranda, 2019), although the idea that physiological traits are responsible for BCI illiteracy is controverted (Thompson, 2019; Riquelme-Ros et al., 2020). These limitations impact replay and the difficulty of games, as players who do not succeed in a BCI task will become stuck in the game without knowing how to improve. In fact, except for a proof of concept or a contest, such as the game created by Mentalista (Paris, France), these games are not suitable for public.

DISCUSSION AND CONCLUSION

This section discusses BCI limitations and explains why overcoming these limitations to develop a ubiquitous BCI game is challenging. These difficulties include a plateau in performance compared with mechanical inputs, technical and algorithmic complexity behind BCI, a lack of middleware for BCI development and an underestimation of graphics and design complexity compared with games in the market.

Transfer rate is the most common limitation discussed in the literature. For example, Rashid et al. (2020) has argued that "most BCI games demonstrate very low accuracy and speed as compared to conventional interfaces, suggesting that there are issues that must be addressed to facilitate the acceptance of BCI games." According to Cattan et al. (2018b) and Rashid et al. (2020), this is particularly true for games requiring movements, such as VR games, as muscular artifacts interfere with the detection of brain signals. In this regard, the complexity of signal detection and classification is, to our belief, a key obstacle to creating effective BCIs for use in games.

From our perspective, recent developments in non-metric (e.g., Quemy, 2019) or quantum classification (Grant et al., 2018; Havenstein et al., 2019) *might* lead to significant improvements in BCI acceptance and performance. Indeed, non-metric classification reduces the need for data pre-processing and engineering, while quantum classification takes advantage of quantum physics to improve the speed and accuracy of classification. The emerging field of quantum machine learning has become increasingly mature thanks to the availability of open-source toolkits (e.g., Abraham et al., 2019) and cloud-based quantum machines (such as the IBM quantum experience by IBM, Armonk, US). Havenstein et al. (2019) showed the advantages of using a quantum vs. classical support vector machine for multi-class classification. Further interesting developments are expected in this field in the next decade.

Nevertheless, in our opinion, the impact of a low-transfer rate on BCI games is overestimated. In practice, if BCIs are considered an interesting yet dispensable device for video games, this is mostly due to design issues because BCIs are either used as an ancillary feature of games (despite requiring expensive materials) or as a means of competing with traditional inputs (e.g., keyboards and mice) to achieve the same task faster and with less concentration. In other words, BCIs in games should be limited to a set of aspects that cannot be achieved by traditional inputs, but at the same time should create sufficient value to justify the cost and encumberment of the material. However, despite some of the positive features previously enumerated (e.g., the use of BCIs for naturally imprecise behavior), a complete game concept that can be sold for concrete video game entertainment is lacking. Further, design reflection for BCI games is still in its infancy and is close to the prototypal-use cases created in the early 2000s (e.g., Bayliss and Ballard, 2000). Similar concerns were broached in recent studies, such as Vasiljevic and Miranda (2019) which reported that only a few studies focused on qualitative aspects of the interaction with BCI games.

In short, developers should above all be concerned with game design and game portability. In practice, BCI games cannot be downloaded and run independently of the EEG acquisition system, which is an impediment for both researchers and game developers. In this respect, the work achieved by platforms such as OpenVibe (Renard et al., 2010) or open-source initiatives that rely on a standard protocol, such as Lab Streaming Layer (Stenner et al., 2015), should be emphasized (see text footnote 7). Nevertheless, there is no support for developers (at any level, technologic, mathematical, usability) to integrate BCI in concrete game production. To our knowledge, practical obstacles such as compilation and integration of LSL DLL into game engines, along with the lack of command-line support or synchronization solutions between OpenVibe acquisition server and recreational applications, are rarely mentioned in the existing literature.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Abraham, H., AduOffei, Agarwal, R., Akhalwaya, I. Y., Aleksandrowicz, G., Alexander, T., et al. (2019). *Qiskit: An Open-source Framework for Quantum Computing*. doi: 10.5281/zenodo.2562110

Ahn, M., Lee, M., Choi, J., and Jun, S. (2014). A review of brain-computer interface games and an opinion survey from researchers, developers and users. *Sensors* 14, 14601–14633. doi: 10.3390/s140814601

Allison, B., and Neuper, C. (2010). "Could anyone use a BCI?," in *Brain-Computer Interfaces: Human-Computer Interaction Series,* eds D. Tan and A. Nijholt (London: Springer), 35–54. doi: 10.1007/978-1-84996-272-8_3

Andreev, A., Barachant, A., Lotte, F., and Congedo, M. (2016). *Recreational Applications of OpenViBE: Brain Invaders and Use-the-Force*. Hoboken, NJ: John Wiley; Sons. Available online at: https://hal.archives-ouvertes.fr/hal-01366873/document (accessed December 11, 2020).

Andreev, A., Cattan, G., and Congedo, M. (2019). *Engineering Study on the Use of Head-Mounted Display for Brain-Computer Interface*. GIPSA-Lab. Available online at: https://hal.archives-ouvertes.fr/hal-02166844 (accessed August 2, 2019).

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., and Congedo, M. (2014). A Plug&Play P300 BCI using information geometry. *ArXiv14090107 Cs Stat*. Available online at: http://arxiv.org/abs/1409.0107 (accessed December 11, 2020).

Bayliss, J. D., and Ballard, D. H. (2000). A virtual reality testbed for brain-computer interface research. *IEEE Trans. Rehabil. Eng. Publ. IEEE Eng. Med. Biol. Soc.* 8, 188–190. doi: 10.1109/86.847811

Cattan, G., Andreev, A., Maureille, B., and Congedo, M. (2018a). *Analysis of Tagging Latency When Comparing Event-Related Potentials*. Grenoble: Gipsa-Lab; IHMTEK. Available online at: https://hal.archives-ouvertes.fr/hal-01947551 (accessed December 11, 2020).

Cattan, G., Mendoza, C., Andreev, A., and Congedo, M. (2018b). Recommendations for integrating a P300-based brain computer interface in virtual reality environments for gaming. *Computers* 7:34. doi: 10.3390/computers7020034

Cattan, G. H., Andreev, A., Mendoza, C., and Congedo, M. (2019). A comparison of mobile VR display running on an ordinary smartphone with standard PC display for P300-BCI stimulus presentation. *IEEE Trans. Games* 1. doi: 10.1109/TG.2019.2957963. [Epub ahead of print].

Chabance, M., Cattan, G., and Maureille, B. (2019). *Implementation of a Daemon for OpenBCI*. IHMTEK. Available online at: https://hal.archives-ouvertes.fr/hal-02091644 (accessed August 3, 2019).

Congedo, M. (2013). *EEG Source Analysis*. Available online at: https://tel.archives-ouvertes.fr/tel-00880483 (accessed December 11, 2020).

Debener, S., Minow, F., Emkes, R., Gandras, K., and de Vos, M. (2012). How about taking a low-cost, small, and wireless EEG for a walk? *Psychophysiology* 49, 1617–1621. doi: 10.1111/j.1469-8986.2012.01471.x

Ferreira, A. L. S., de Miranda, L. C., de Miranda, E. E. C., and Sakamoto, S. G. (2013). A survey of interactive systems based on brain-computer interfaces. *J. Interact. Syst.* 4, 3–13. doi: 10.5753/jis.2013.623

Grant, E., Benedetti, M., Cao, S., Hallam, A., Lockhart, J., Stojevic, V., et al. (2018). Hierarchical quantum classifiers. *Npj Quantum Inf.* 4, 1–8. doi: 10.1038/s41534-018-0116-9

Havenstein, C., Thomas, D., and Chandrasekaran, S. (2019). *Comparisons of Performance between Quantum and Classical Machine Learning*. SMU Data Science Review, 1. Available online at: https://scholar.smu.edu/datasciencereview/vol1/iss4/11 (accessed December 11, 2020).

Kaplan, A. Y., Shishkin, S. L., Ganin, I. P., Basyul, I. A., and Zhigalov, A. Y. (2013). Adapting the P300-based brain-computer interface for gaming: a review. *IEEE Trans. Comput. Intell. AI Games* 5, 141–149. doi: 10.1109/TCIAIG.2012.2237517

Kerous, B., Škola, F., and Liarokapis, F. (2018). EEG-based BCI and video games: a progress report. *Virtual Real.* 22, 1–17. doi: 10.1007/s10055-017-0328-x

Lotte, F., Renard, Y., and Lécuyer, A. (2008). Self-Paced brain-computer interaction with virtual worlds: a quantitative and qualitative study "Out of the Lab," in *4th international Brain Computer Interface Workshop and Training Course* (Graz: Graz University of Technology). Available online at: https://hal.inria.fr/inria-00304340 (accessed December 11, 2020).

Marshall, D., Coyle, D., Wilson, S., and Callaghan, M. (2013). Games, gameplay, and BCI: the state of the art. *IEEE Trans. Comput. Intell. AI Games* 5, 82–99. doi: 10.1109/TCIAIG.2013.2263555

Nijholt, A., Plass-Oude Bos, D., and Reuderink, B. (2009). Turning shortcomings into challenges: brain-computer interfaces for games. *Entertain. Comput.* 1, 85–94. doi: 10.1016/j.entcom.2009.09.007

Pierce, S., S, C. C., Marvin, A., Anton, N., and E, G. J. (2020). Brain-computer interface software: a review and discussion. *IEEE Trans. Hum. Mach. Syst.* 50, 101–115. doi: 10.1109/THMS.2020.2968411

Quemy, A. (2019). Binary classification in unstructured space with hypergraph case-based reasoning. *ArXiv180606232 Cs Stat*. Available online at: http://arxiv.org/abs/1806.06232 (accessed August 29, 2020).

Rashid, M., Sulaiman, N., Abdul Majeed, A. P. P., Musa, R. M., Ab. Nasir, A. F., Bari, B. S., et al. (2020). Current status, challenges, and possible solutions of EEG-based brain-computer interface: a comprehensive review. *Front. Neurorobotics* 14:25. doi: 10.3389/fnbot.2020.00025

Renard, Y., Lotte, F., Gibert, G., Congedo, M., Maby, E., Delannoy, V., et al. (2010). OpenViBE: an open-source software platform to design, test, and use brain–computer interfaces in real and virtual environments. *Presence Teleoperators Virtual Environ.* 19, 35–53. doi: 10.1162/pres.19.1.35

Riquelme-Ros, J.-V., Rodríguez-Bermúdez, G., Rodríguez-Rodríguez, I., Rodríguez, J.-V., and Molina-García-Pardo, J.-M. (2020). On the better performance of pianists with motor imagery-based brain-computer interface systems. *Sensors* 20:4452. doi: 10.3390/s20164452

Rosca, S., and Leba, M. (2019). Design of a brain-controlled video game based on a BCI system. *MATEC Web Conf.* 290:01019. doi: 10.1051/matecconf/201929001019

Schell, J. (2014). *The Art of Game Design: A Book of Lenses, Second Edition*. Natick, MA: A K Peters (CRC Press).

Stenner, T., Boulay, C., and Medine, D. (2015). *LabStreamingLayer*. Swartz Center for Computational Neuroscience. Available online at: https://github.com/sccn/labstreaminglayer (accessed December 11, 2020).

Thompson, M. C. (2019). Critiquing the concept of BCI illiteracy. *Sci. Eng. Ethics* 25, 1217–1233. doi: 10.1007/s11948-018-0061-1

van de Laar, B., Gürkök, H., Bos, D. P.-O., Poel, M., and Nijholt, A. (2013). Experiencing BCI control in a popular computer game. *IEEE Trans. Comput. Intell. AI Games* 5, 176–184. doi: 10.1109/TCIAIG.2013.2253778

Vasiljevic, G., and Miranda, L. (2019). Brain-computer interface games based on consumer-grade EEG devices: a systematic literature review. *Int. J. Hum. Comput. Interact.* 36, 105–142. doi: 10.1080/10447318.2019.1612213

Vos, M. D., Kroesen, M., Emkes, R., and Debener, S. (2014). P300 speller BCI with a mobile EEG system: comparison to a traditional amplifier. *J. Neural Eng.* 11:036008. doi: 10.1088/1741-2560/11/3/036008

Wolpaw, J., and Wolpaw, E. W. (2012). *Brain-Computer Interfaces: Principles and Practice*. New York, NY: Oxford University Press.

Yohanandan, S. A. C., Kiral-Kornek, I., Tang, J, Mshford, B. S., Asif, U., and Harrer, H. (2018). A robust low-cost EEG motor imagery-based brain-computer interface. *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2018, 5089–5092. doi: 10.1109/EMBC.2018.8513429

Conflict of Interest: GC was employed by the company IBM.

frontiers in Computer Science

ORIGINAL RESEARCH
published: 14 April 2021
doi: 10.3389/fcomp.2021.661300

Check for updates

Brain-Computer Interfaces, Open-Source, and Democratizing the Future of Augmented Consciousness

Guillermo Bernal[1†], Sean M. Montgomery[2†] and Pattie Maes[1]*

[1] Media Laboratory, Massachusetts Institute of Technology, Cambridge, MA, United States, [2] Connected Future Labs, New York, NY, United States

Accessibility, adaptability, and transparency of Brain-Computer Interface (BCI) tools and the data they collect will likely impact how we collectively navigate a new digital age. This discussion reviews some of the diverse and transdisciplinary applications of BCI technology and draws speculative inferences about the ways in which BCI tools, combined with machine learning (ML) algorithms may shape the future. BCIs come with substantial ethical and risk considerations, and it is argued that open source principles may help us navigate complex dilemmas by encouraging experimentation and making developments public as we build safeguards into this new paradigm. Bringing open-source principles of adaptability and transparency to BCI tools can help democratize the technology, permitting more voices to contribute to the conversation of what a BCI-driven future should look like. Open-source BCI tools and access to raw data, in contrast to black-box algorithms and limited access to summary data, are critical facets enabling artists, DIYers, researchers and other domain experts to participate in the conversation about how to study and augment human consciousness. Looking forward to a future in which augmented and virtual reality become integral parts of daily life, BCIs will likely play an increasingly important role in creating closed-loop feedback for generative content. Brain-computer interfaces are uniquely situated to provide artificial intelligence (AI) algorithms the necessary data for determining the decoding and timing of content delivery. The extent to which these algorithms are open-source may be critical to examine them for integrity, implicit bias, and conflicts of interest.

Keywords: augmentation, art, closed-loop systems, virtual realities, open-source, brain computer interface

OPEN ACCESS

Edited by:
Anton Nijholt,
University of Twente, Netherlands

Reviewed by:
Elizabeth A. Boyle,
University of the West of Scotland, United Kingdom
Dulce Fernandes Mota,
Instituto Superior de Engenharia do Porto (ISEP), Portugal

****Correspondence:***
Guillermo Bernal
gbernal@mit.edu

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Computer Science

Received: *30 January 2021*
Accepted: *16 March 2021*
Published: *14 April 2021*

Citation:
Bernal G, Montgomery SM and Maes P (2021) Brain-Computer Interfaces, Open-Source, and Democratizing the Future of Augmented Consciousness. Front. Comput. Sci. 3:661300. doi: 10.3389/fcomp.2021.661300

INTRODUCTION

Brain-computer interfaces (BCIs) are poised to transform the nature of human consciousness in the 21st century. In this context, we adopt the operational definition of being conscious as having an experience – the subjective phenomena "what it's like" to see an image, hear a sound, conceive a thought, or be aware of an emotion (Sandberg et al., 2010; Faivre et al., 2015; Koch et al., 2016). We speculate based on prior research that closed-loop systems with a combination of stimuli (mixed reality), sensing (BCI) and predictive algorithms (AI and its subsets machine learning and deep learning) will increasingly be capable to alter the subjective experience in a manner that is tightly coupled with changes in emotion regulation (Lorenzetti et al., 2018; Montana et al., 2020), and cognitive augmentations such as attention improvements (Wang et al., 2019), episodic memory enhancement (Burke et al., 2015). These transformations may not only make humanity more productive and efficient, but also potentially more expressive, understanding, and empathetic.

We can begin by classifying BCIs into three main groups: (1) invasive approach (Wolpaw et al., 2000), (2) partially invasive approach, and (3) non-invasive approach. An invasive approach requires electrodes to be physically implanted into the brain's gray matter by neurosurgery, making it possible to measure local field potentials. A partially invasive approach (e.g., electrocorticography—ECoG) is applied to the inside of the skull yet outside the gray matter. A non-invasive approach (e.g., electroencephalography—EEG and functional Magnetic Resonance Imaging—fMRI) is the most frequently used signal capturing method. This system is placed outside of the skull on the scalp and records the brain activities inside of the skull and on the surface of the brain membranes. Both EEG and fMRI give different perspectives and enable us to "look" inside the brain (Kropotov, 2010). It is worth noting that invasive and partially invasive approaches are prone to scar tissue, are difficult to operate, and expensive. Although EEG signals can be prone to noise and signal distortion, they are easily measured and have a good temporal resolution. This and the fact that fMRI devices are expensive and cumbersome to operate make EEG the most widely used method for recording brain activity in BCI systems. EEG-based devices directly measure electrical potentials produced by the brain's neural synaptic activities.

While the use of EEG was originally limited to medical imaging research labs, more compact and affordable EEG systems have opened up opportunities for other applications to be explored. In this paper, we discuss some of the ways that these new technologies have been applied in areas ranging from the arts, self-improvement and rehabilitation to gaming and augmented reality. While consumer electronics devices have increased the accessibility to BCI technologies, we also discuss some of the ways in which these same devices limit their adaptability beyond pre-defined use-cases as well as the transparency of the data and algorithms. For the context of this discussion, we use BCI technology to describe devices that measure a broad array of biometric signals, not only directly from the central nervous system (CNS), but also from the peripheral nervous system (PNS). Because changes in cognitive and emotional states engage sympathetic and parasympathetic responses of the PNS, changes in heart rate, electrodermal activity, and other biometric signals can provide a detailed window into brain activity (Picard, 1995).

Our discussion briefly reviews the evolution of BCI devices with examples of how they have been applied outside of traditional research settings. Within a transdisciplinary context, including neuroscience, computer science, health, philosophy, art, and a rapidly developing technology landscape, we review specific ways in which limitations to the adaptability and transparency of BCI technology can have implications for applications both within and outside research contexts. We examine how applying open-source principles may help to democratize the technology and overcome some of these limitations, both for traditional research as well as alternative uses. Looking forward to a future in which BCIs are likely to become increasingly integrated into our everyday lives, we believe that it will be important to involve more voices from across traditional disciplinary divides contributing to the conversation about what our future should look like and how BCI devices and data should contribute to our lives. From public BCI art to hackathons and K-12 education, it will likely be critical to be asking more questions, new types of questions from different perspectives, and starting at a younger age, to ensure that BCI technology will serve society at large.

As BCIs move beyond siloed research labs toward new and more diverse use cases, the accessibility, adaptability, and transparency of BCI tools and data will significantly impact how we collectively navigate this new digital age. Technology and the self are becoming increasingly coupled, allowing us to learn faster, create new ways to express ourselves and share information like never before. The extent to which these technologies are open and accessible for more people to engage with them and examine their integrity may shape the nature of our consciousness and the future of humanity.

CHANGE IN PERSPECTIVE

Artists and do-it-yourself-ers (DIYers) have been exploring novel BCI applications since before BCI was an acronym. Artists and DIYers often adopt new technologies to modify their original condition and purpose beyond the "intended" use. Going back 60 years, artists proposed novel experiments such as the sonification of alpha waves to excavate untapped musical imaginations or subconscious musicality. There was especially a great interest by composers in the use of feedback, both acoustic and electronic, as a fundamental musical process. In 1965, physicist Edmond Dewan and composer Alvin Lucier collaborated on Music for the Solo Performer, shown in **Figure 1**. This piece is generally considered to be the first musical work to use brain waves and directly translate them into musical sound. Lucier's work remains a pioneering and important piece of 20th-century music, as well as one of the touchstones of early "live" electronic music. Classic feedback pieces such as David Behrman's WaveTrain (Behrman, 1998), Max Neuhaus' Public Supply (Max, 1977), and Terry Riley's tape delay feedback (Sitsky, 2002) were also created in this new wave of exploration. These early artists were often found

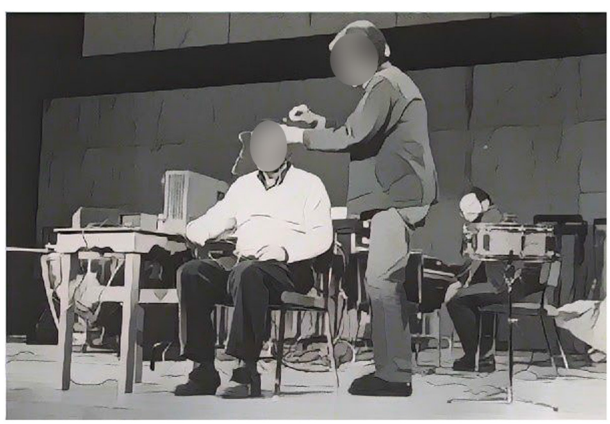


FIGURE 1 | Physicist Edmond Dewan and composer Alvin Lucier collaborated on music for the solo performer.

building BCIs and synthesizers from scratch using basic electrical components (resistors, capacitors, amplifiers, etc.), which gave them a very high degree of flexibility to create and adapt their circuits to do strange and wonderful things but also posed a high barrier of entry to create and use the technology.

The explosion of consumer electronics during the 80 and 90's resulted in substantial moves in the BCI space in the late 2000's and early 2010's. NeuroSky, Emotiv, and Zeo Sleep Coach among others developed consumer EEG devices and BodyMedia, Polar, Fitbit and other devices measured signals from the autonomic nervous system (Nijboer et al., 2011). This increase in accessibility of BCI technologies has enjoyed a commensurate blossoming of BCI-driven art. Emotiv's EPOC headset has been utilized in a number of artistic contexts including measuring the Magic of Mutual Gaze (Abramovic et al., 2014), a durational performance art piece by Marina Abramovic that utilized the Emotiv EPOC to visualize and neuro-contextualize the synchrony of two people engaged in mutual gaze (as shown in **Figure 2**). Noor: A Brain Opera (Pearlman, 2016) turned brainwave data into an immersive audio-visual experience in which the internal state of the performers drives the operatic performance. Vessels (Leslie, 2020) is a brain-body performance piece that combines flute improvisation with live, sonified brain and body data from a Muse headband. Polar handheld devices similarly became the springboard for a number of artworks including Pulse Spiral (Lozano-Hemmer, 2008) and Emergence, as shown in **Figure 3** (Montgomery et al., 2011), measuring viewer's heartbeats to reflect on the nature of their internal state and the human-computer interface.

While this consumer electronics explosion made BCI technology more accessible, the resulting devices (including the combination of hardware, firmware, and required software) often place limits on the adaptability and extensibility of the device. In many cases raw data is made completely inaccessible or is only offered at high price points, tending to confine the resulting artistic applications into more pre-defined use-cases.

Around the same time as the advent of consumer-grade BCI devices, the DIY maker movement began to take root in the 2000's, in part to break down the dichotomy between accessibility and flexibility/extensibility. One of the movement's cornerstones was born when a group of Italian postgraduate students and a lecturer at the Interaction Design Institute in Ivrea created the first version of what would go on to become the Arduino project (Banzi et al., 2015). The OpenEEG project (Griffith, 2006) quickly became an early go-to open-source circuit for everything from academic tutorials and clothing that lights up with brain activity to increase the expressiveness of the wearer (Montgomery, 2010) to adaptive drone piloting (Ossmann et al., 2014), but the volume-pricing of consumer electronics quickly became an attractive opportunity for artists and DIYers alike (Montgomery and Laefsky, 2011) to find "alternative uses" for the devices. Hacking of the Mattel MindFlex (How to Hack Toy EEGs | Frontier Nerds, 2010) to extract derivative EEG power-band data and of the Zeo Sleep Coach to extract raw EEG data (Dan, 2011) enabled a multiplicity of art installations such as Telephone Rewired, using rhythmic visual and audio patterns to alter endogenous brain oscillations and create an immersive aesthetic experience and altered subjective state (Produce Consume Robot and LoVid, 2013) and Teletron by the band Apples in Stereo, an instrument which allows the user to play an analog synth completely through brain activity (Schneider, 2010). These projects leveraged the wearability of consumer-grade EEG devices to break down barriers between artistic expression and scientific research.

In the 2010's, Pulse Sensor, OpenBCI, EmotiBit, and other fully open-source products with full data access further broke down the barriers to accessibility and extensibility that gave artists and creative technologists as well as researchers and educators access to high-quality BCI platforms beyond the confines of specific intended uses (Hoffman and Bast, 2017; Montgomery, 2018a; Gupta et al., 2020; Masui et al., 2020; Vujic et al., 2020). Ever since, developers have been fascinated with

FIGURE 2 | Measuring the magic of mutual gaze & the artist is present. Left: Measuring the magic of mutual gaze at the garage museum for contemporary art, Moscow in 2011. Photograph by Maxim Lubimov © Garage center for contemporary culture. Video: www.youtube.com/watch?v=Ut9oPo8sLJw&t=73s Right: Marina Abramović, the artist is present, performance, 3 months, the museum of modern art, New York, NY (2010), photography by Marco Anelli. Courtesy of the Marina Abramović Archives.

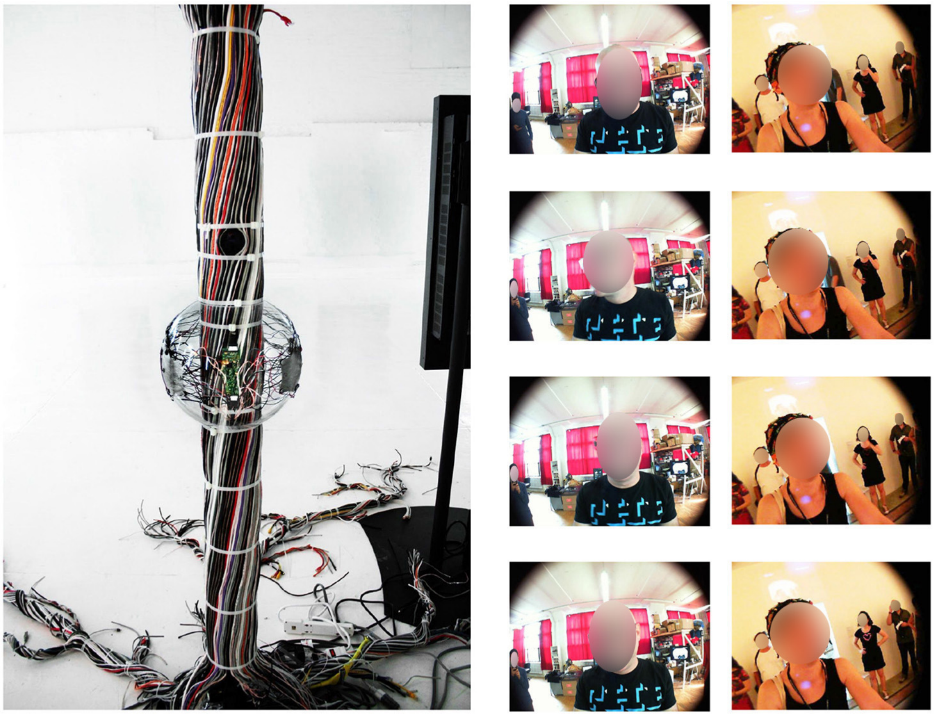


FIGURE 3 | Emergence installation at open house gallery, New York City (left) and digital memories triggered by successive heartbeats and uploaded to Flickr (right). Left image by Sean Montgomery. Right images by Emergence courtesy of Sean Montgomery. Web: http://produceconsumerobot.com/emergence/; Leonardo Electronic Almanac Vol 18 No 5 p 6–9.

the possibility of enhancing the gaming experience via BCIs (Lécuyer et al., 2008). Games tailored to the user's affective state—immersion, flow, frustration, surprise, etc.—like the famous World of Warcraft, allow an avatar's appearance to reflect the gamer's cognitive state instead of being controlled through keyboarding (Nijholt et al., 2009). It is not unrealistic to believe that the first mass application of non-medical BCIs will be in the gaming and entertainment field. Standalone examples already have a market, and extensions to console games are likely to follow soon (Nijboer et al., 2011). Other projects like Emotional Beasts allowed the exploration of the user's self-expression in VR space by transforming the appearance of the avatar in an artistic way based on the user's affective state, thereby pulling the avatar design away from the uncanny valley problem and making it more expressive and more relatable (Bernal and Maes, 2017). Through the use of VR headsets that have been altered to accommodate physiological sensors (Bernal et al., 2018), the system collects and integrates physiological data to enable the perception of human affect. Bernal et al. showed how the PhysioHMD system can be used to help develop personalized phobia treatment by creating a closed loop experience. The images (insect sprites) spawned through the particle system can be modified (speed, size, the rate of spawn, movement) in the Unity inspector to increase or decrease the arousal level of the user.

As we look forward to the 2020's and beyond, it's reasonable to expect that BCI technology will become a greater part of everyday life for humanity and that these technologies may integrate with and potentially change aspects of human cognition. The extent to which artists and makers are enabled to be a part of the conversation about what this future should look like and where there are potential dystopian hazards, may play a pivotal role in shaping that future (Flisfeder, 2018; Montgomery, 2018b). The level of engagement and dialog will depend on the accessibility made possible through volume production of consumer electronics, the adaptability and flexibility made possible through open-source technology, and granular access to raw data that allows for going beyond pre-baked intended uses to ask new questions about brains, computers and the interfaces that increasingly connect them.

MY DATA, THEIR DATA?

Looking Behind the Curtain

As availability of consumer-grade BCI and biometric technologies has grown, the application of these technologies in research, serious games, and rehabilitation has surged (View Research, 2019). Instead of costing tens of thousands of U.S. dollars and requiring participants to wear saline-soaked headgear entangled with dozens of wires, it is now possible to get meaningful BCI data from devices costing < US$1,000 in wireless and relatively easy to wear form factors, as shown in **Figure 4**. The result has vastly opened up possibilities for more people around the world to ask more and new types of

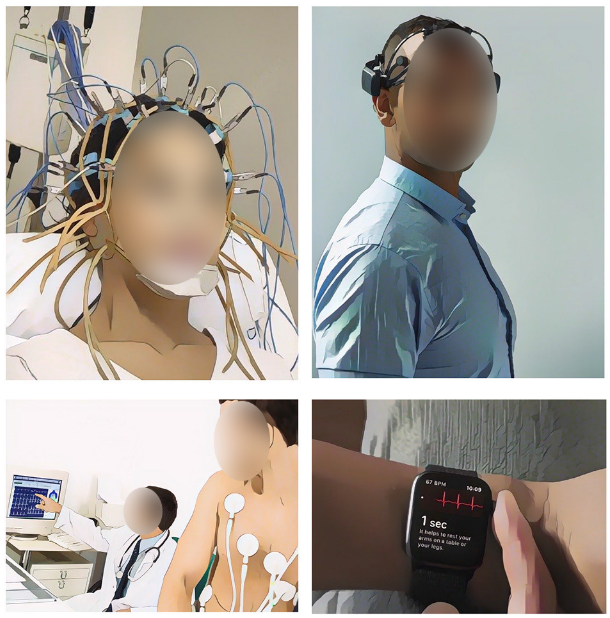


FIGURE 4 | Consumer-grade BCI devices substantially improved the comfort and wearability, permitting devices to be used in new contexts and new types of questions to be asked. Top left, a standard wet EEG system, earlier systems were very susceptible to movement artifacts. Top right, a wireless headset by Emotiv, still uses wet electrodes but the setup is much quicker and wearability much improved compared to earlier systems. Bottom left, ECG setup found in most clinics, the setups commonly use 10 sticky electrodes to monitor cardiac activity. Bottom right, today's wrist-worn smartwatches can report user's ECG from the press of a button.

questions. For rehabilitation, that means it is easier to bring devices into people's homes or care facilities, allowing for many more people to be served and for exploring new treatments and methodologies (Sung et al., 2012). Bringing consumer BCI technologies into the workplace has led to rapid development ranging from serious games, to flight simulators and warehouse safety training (Marshall et al., 2013; Huang et al., 2020) and research labs can now ask new questions about the neuroscience of interpersonal and classroom dynamics (Dikker et al., 2017, 2019a).

With these benefits of consumer-grade BCI devices, however, has also come a challenge of transparency. Most of the consumer-electronics world, including BCI devices, lives and dies by a closed-source ethos. While patents can serve as a mechanism to maintain transparency as well as a competitive advantage, it is often difficult or impossible to patent BCI circuits and algorithms because they are either not sufficiently novel or can be modified slightly to avoid infringement claims, and yet, omitting specific details (e.g., filter frequencies and other signal conditioning) can both make it harder for others to simply rip off the product as well as limit the ways the end-user can utilize and interpret the data. The patent system is also rather slow and somewhat overrun, so in a fast-moving technology landscape the competitive advantage of a patent might be somewhat irrelevant by the time a patent is actually granted. In addition to limiting access to raw data and only providing end-users with summary statistics, a competitive advantage is commonly maintained by creating a moat of closed-source trade-secrets and datasets that prevent competitors from moving into the space. On the other hand, essential to the world of science is the principle of reproducibility, and in a number of areas the scientific method and the closed-source veil stand on different sides of the table propelling the stalwart march of human knowledge (McNutt, 2014; Höller et al., 2017).

One way closed-source/closed-data ecosystems limit research is in the scope of questions that are possible to ask. For example, when examining EEG data, if only power-band statistics (alpha, beta, gamma, etc.) are made available (as for example is the case for the standard Emotiv license), much of the information about synchrony in the brain is lost. Specifically, it is impossible to investigate whether two regions of the brain are exhibiting coherence or phase-locking with one another. There is wide consensus among neuroscientists (Uhlhaas et al., 2009) that synchrony is important in determining the efficacy of neuronal communication, plasticity, and learning, and possibly even for governing aspects of consciousness (Buzsaki, 2006). For example, the phase-locking of EEG oscillations has been shown to increase between different regions of the medial-temporal lobe during successful memory formation (Miltner et al., 1999; Fries, 2015) suggesting an important role in memory encoding or selective attention (Fries, 2015). Similarly, increased coherence has been associated with memory retrieval (Kaplan et al., 2014; Meyer et al., 2015) and binding together of multi-modal representations spanning different areas of the brain (Gray et al., 1989; Llinás et al., 1998; Engel et al., 2001). In the context of this research, it is likely that timing and synchrony in the brain are important for some of the most interesting cognitive functions—memory encoding and retrieval, associative learning, attention, and likely others. However, when EEG data is reduced to power-band statistics, the phase relationships and cycle-by-cycle timing information is inherently lost. In doing so, it is possible that some of the most important information about the operation of the brain and how it relates to cognition may be lost in an irreversible way. Similarly, as we increasingly apply machine learning to EEG data, if only power-band data or other potentially impoverished derivative metrics are used as inputs, this may fundamentally limit the effectiveness of the resulting machine learning models. In some cases the resulting models may lack statistical power to make reliable predictions and in other cases the models may overfit the power-band data and be unable to adequately generalize and replicate the results in other contexts. As the field of neuroscience continues working to understand what the most important parameters of brain function are that derive cognitive processes, having access to the raw EEG data will likely be critical to unlock the true potential of this technology.

Further limitations can come when deriving metrics from data preprocessed with closed-source algorithms. For example, using heart rate (HR) data from common wrist-worn devices (e.g., FitBit or Apple Watch) to calculate pulse rate variability (PRV) can be substantially problematic. Similar to heart rate variability (HRV), PRV is a metric that is particularly dialed into the sympathetic/parasympathetic axis of the autonomic nervous system (Berntson et al., 1997). By looking at the relationship between fast and slow changes in HR it is possible to mathematically relate heart rate changes to levels of epinephrine governing sympathetic "fight or flight" responses and levels of acetylcholine governing parasympathetic "rest and digest" responses (Appelhans and Luecken, 2006). However, PRV calculations require long periods (typically 10+ min) of clean data, to accurately estimate the low frequency (LF) and high frequency (HF) components used to assess the sympathetic/parasympathetic ratio. While measuring photoplethysmography (PPG) from the wrist is convenient, one of the limitations is that it can be subject to substantial movement artifacts depending on a host of factors (Biswas et al., 2019). To deal with movement-related noise, consumer-grade devices typically employ heavy-handed smoothing and interpolation in order to give consumers a "best guess" HR value even when the signal quality is low. While this interpolation can improve the HR estimation and the consumer experience overall, as is illustrated in **Figure 5**, the resulting PPG estimation can be very substantially distorted (Morelli et al., 2018). Without access to the original raw PPG data, researchers are simply stuck with the HR estimates coming off the device with no path by which to improve the HR detection as new algorithms are developed, for example, based on recent research using a sensor fusion approach combining PPG and accelerometer data (Kos et al., 2017; Biswas et al., 2019). In contrast, open-source products

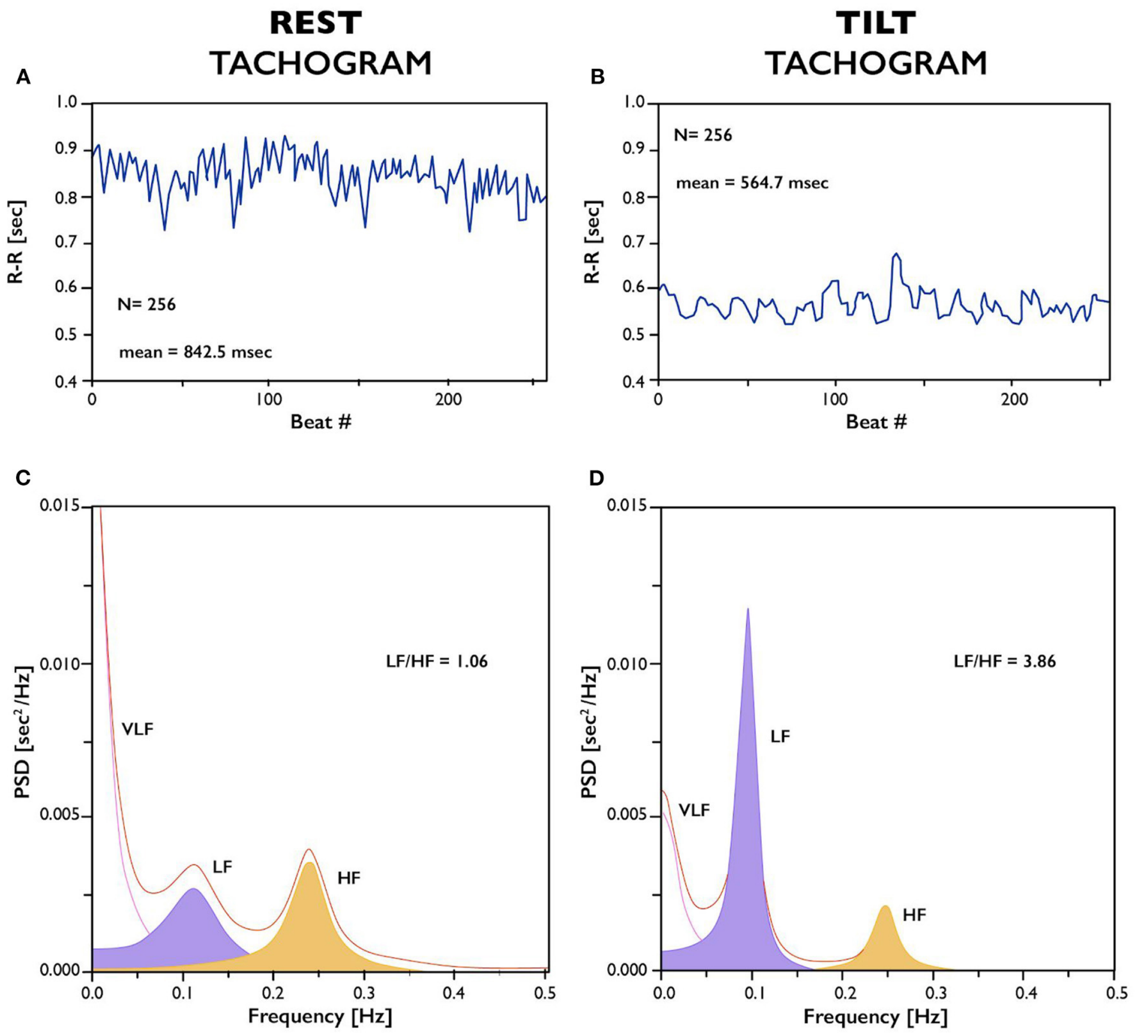


FIGURE 5 | Illustrative example of heart rate variability derivation and how smoothing heart rate data can lead to variability detecting changes in the sympathetic/parasympathetic nervous system. Figure plots are based on data presented in Electrophysiology (1996). **(A)** shows the raw tachogram fluctuations in heart rate during supine rest and **(C)** shows the derivative power spectral density (PSD) of the supine rest heart rate data to calculate the VLF, LF, and HF frequency bands that can be used to assess autonomic nervous system balance. **(B,D)** show the raw heart rate tachogram and derived PSD plot after a 90 degree head-up tilt physiological perturbation that increases the sympathetic nervous system response. Smoothing or interpolation algorithms acting on the raw rest tachogram data can potentially generate tachogram data similar to the tilt condition in **(B)**, leading to a spurious shift in the observed LF/HF ratio. In the context of wearable consumer devices with potential data gaps and heavy-handed closed-source smoothing/interpolation algorithms, it is thereby possible to misinterpret smoothed or interpolated data as a shift in the sympathetic/parasympathetic nervous system responding, even when no such shift has occurred.

like OpenBCI (Murphy and Russomanno, 2013), Pulse Sensor (WorldFamousElectronicst, 2011) and EmotiBit (https://www.emotibit.com/, Montgomery 2020) provide access to the raw data as well as the electrical specifications and source code to understand the data collection and derivatives.

Worse than the data distortion itself is that the algorithms performing the interpolation on consumer-grade devices are closed-source and it is often unclear when data is being interpolated and when it is faithfully reflecting the physiological activity of the wearer. As a result, it can be very difficult to assess when the data distortion may be leading to interpretations that are overstated, understated, or even opposite of the truth in any given experimental paradigm. Potentially even more problematic for the use in scientific contexts is that the algorithms deriving biometrics can change without notice any time the company chooses to push new code onto the device, phone or remote servers. These quickly evolving algorithms are especially difficult to reveal and protected with patents, and algorithm changes with different firmware and software releases are thus always a potential caveat when a result fails to replicate from one study to the next or even if an effect shows up, disappears, or changes in the midst of a single study (Shcherbina et al., 2017). When it comes to developing new therapeutic approaches, training protocols and serious game applications, these kinds of errors can have real-world consequences that can potentially affect people's lives.

When technological tools are developed out in the open, anyone can verify if a vendor is actively pursuing accurate validation metrics, appropriately managing security and privacy, or handling issues in a timely and professional manner. The ability to examine the process followed and the source code developed makes it so that anyone can perform an independent audit. This is true not only for the code itself, but also the methods and testing processes used in the development and the history of changes. The transparency of open-source tools and the access to raw data is similarly important for BCI applications in research as it is for gaming, therapy, and rehabilitation. As state-of-the-art technology endeavors to make sense of signals from the body, it is often critical to understand important details of how the data was collected, conditioned, and transformed into derivative metrics. Particularly as derivative metrics become building blocks and inputs for downstream analysis and machine learning algorithms, transparency becomes essential for the ability to replicate and understand results and unlock the mysterious inner workings of the human brain.

Looking Into the Black Box

Stepping forward from simple derivative metrics, the recent wave of machine learning commonly utilizes deep neural network models, which often behave like black boxes because the relationship between the input and output can be difficult to ascertain. When we look into the number of journal articles about ML in neuroscience, we find that its adoption has continuously grown over the last 30 years (**Figure 6**). This rise occurred because Neuroscience has experienced a revolution in the volume of data and datasets that researchers are able to gather from a large number of neurons that researchers can record from, and the size of the datasets is rising rapidly. Researchers increasingly need machine learning methods to wrangle this data and try to gain insight into it. Deep learning, a subset from ML, has given researchers methods for relating high-dimensional neural data to high-dimensional behavior. In addition to their ability to model complex, intelligent behaviors, Deep neural networks (DNNs) excel at predicting neural responses to novel sensory stimuli with accuracy beyond any other currently available model type. DNNs can have millions of parameters required to capture the domain knowledge needed for successful task performance regression models. These parameters are not meant to capture what features of neural activity relate to what features of behaviors, but rather what features of neural activity display information about the behavior or sensory stimuli. If these models aren't made transparent so that BCI experts can interpret the model's decisions based on key model features, it can be challenging to predict the reliability and transferability to new contexts. Current state-of-the-art performance in multi-class EEG ranges from detection of epilepsy (Acharya et al., 2018), to cognitive-workload recognition (Almogbel et al., 2018), to bullying incident identification within an immersive environment (Baltatzis et al., 2017). However, relatively little work in the field of AI or BCI has been done to analyze the interpretability of such models. We define interpretability as the degree to which humans can consistently predict the model's decision (Kim et al., 2016).

It is in this context that the concepts of explain ability and interpretability have taken on new urgency. They will likely only become more important moving forward as discussions around artificial intelligence, data privacy, and ethics continue. The open-source community's recent efforts to support methods and applications that can lead to better trust in AI systems are already producing results. Two open-source methods are available to the public and kept on GitHub to decompose a neural network's output prediction on specific inputs. LIME (Ribeiro et al., 2016) and SHAP (Lundberg and Lee, 2017) are two projects providing novel techniques that explain a classifier's predictions in an interpretable and reliable manner, by learning an interpretable model locally around the prediction. These techniques produce "visual explanations" for decisions from a large class of Convolution Neural Network-based models, making them more transparent and accessible to a human expert by comparing the amount and degree of overlap between identified inputs.

Leveraging the open-source community can help improve trust by ensuring that any BCI-AI effort meets safe and transparent regulations. The community can include domain experts and set routine checks to the codebase. Beyond transparency into the code alone, as our artificial neural network (ANN) models continue to increase in complexity, having tools and transparency to visualize and understand key relationships of the models will be important in understanding when the data and decisions can be trusted and used in research and real-life applications. A closed approach to sophisticated BCI systems can lead to inadequate feature design choices that are not relevant to the current needs of the community and society. Such features can be harmful to the system; for example, if a medical system's patient diagnostic function has poor accuracy due to lack

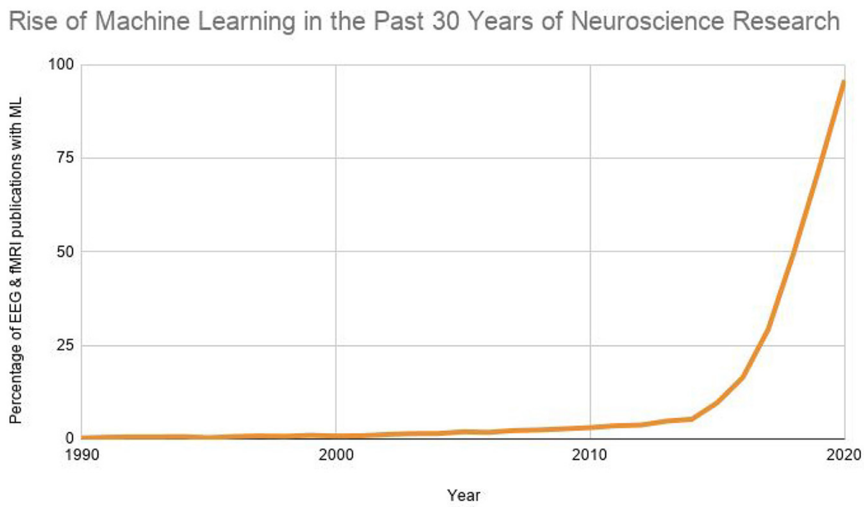


FIGURE 6 | Here we plot the proportion of neuroscience papers that have used machine learning over the last three decades. That is, we calculate the number of papers involving both neuroscience and machine learning, normalized by the total number of neuroscience papers. Neuroscience papers were identified using a search for "EEG + fMRI" on semantic scholar. Papers involving neuroscience and machine learning were identified with a search for "machine learning" and "EEG + fMRI" on semantic scholar.

of testing, then this will mean more human intervention and, ultimately, less trust.

NEW REALITIES AND AUGMENTED COGNITION

Closed Loops

Looking forward to a future in which augmented and virtual reality become an integral part of daily life, BCIs will likely play an increasingly important role in creating closed-loop feedback for generative content. As one physical reality transforms into a multiplicity of mixed realities, brain-computer interfaces will be uniquely situated to provide necessary feature selections to determine the decoding and timing of content delivery.

A closed-loop system that monitors the user's reactions to the content of a virtual environment enables the generation of personalized virtual reality experiences. Demonstrations like those shown by Bernal et al. in **Figure 7** use arousal levels to provide real-time, reliable information about the user's reception of the content and can help the system adapt the content seamlessly (Bernal et al., 2018). In a 360 video demo player scene, they used gaze data and Skin Conductance Response (SCR) data to increase the user's arousal levels by modulating a shader's occlusion superimposed on the 360 scene in order to create the fear of the unknown. **Figure 8** shows how the demo takes standard footage from people in a basement and creates a pulsating shadow effect, and therefore a more dramatic 360 captured video, similar to those seen in horror movies. To direct the user's focus to the people within the video, a surface shader is modulated dynamically to occlude areas that are not of interest to the user; locations informed by the point of regard (POR) data from the gaze tracking system. The detected SCR peak values are used to pulse the occlusion shader with modulation.

Sourina et al. proposed a real-time approach for feature extraction in EEG-enabled applications for serious games, emotional avatars, music therapy, musicians, and storytelling where the emotional states of the users are mapped onto avatars. The Haptek Activex control provides functions and commands to change the six facial expressions of 3D avatars including fear, frustration, sadness, happiness, pleasant, and satisfaction. In the application, emotions of the user are recognized from EEG and visualized in real-time on the user's avatar with the Haptek system. For the music therapy application, the music selection and duration is adjusted based on the current emotional state of the user wearing the BCI, as identified by the system (Sourina et al., 2011).

These types of scenarios don't come without consequences if proper guidelines are not being followed. Recently a BCI start-up has been under scrutiny over tests on Chinese schoolchildren after the Wall Street Journal released a video stating that teachers at that school know exactly when students are and are not paying attention (Wall Street Journal, 2019). In the video, children are shown wearing an EEG headband during class with an LED located on the forehead region that changes color based on the children's attention levels. At the time there were no privacy laws regulating this type of collaboration between private schools and companies. Even though the start-up reports that all parties involved had given consent about taking part in the test, concerns about whether the data was adequately secured or potential future implications for the children led to the ban of this device's use in the classroom and the creation of a new law (Chinese Primary School Halts Trial of Device That Monitors Pupils Brainwaves, 2019; Primary School in China Suspends Use of BrainCo Brainwave Tracking Headband, 2019).

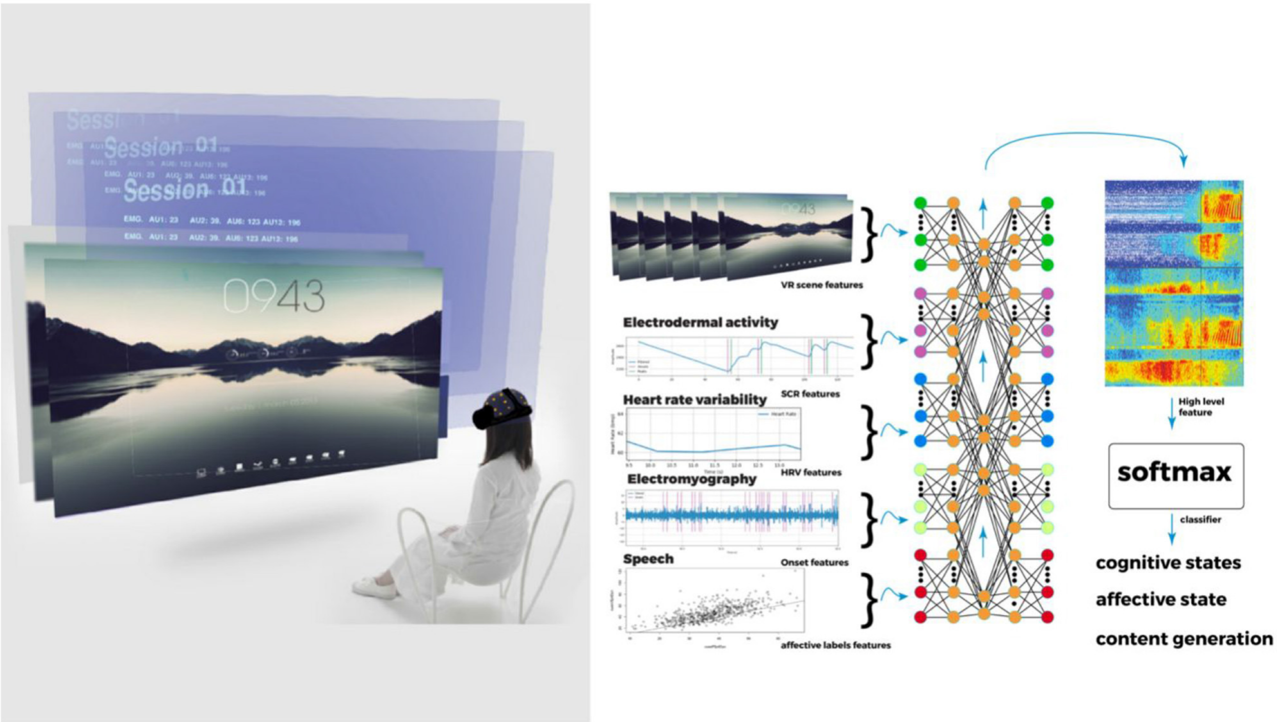


FIGURE 7 | An example of a closed-loop setup used in a lab setting. Physiological signals are recorded using a set of electrodes positioned over the visual cortex while stimuli are displayed on a VR headset. Currently, the system supports deep learning model inference for high-level cognitive states.

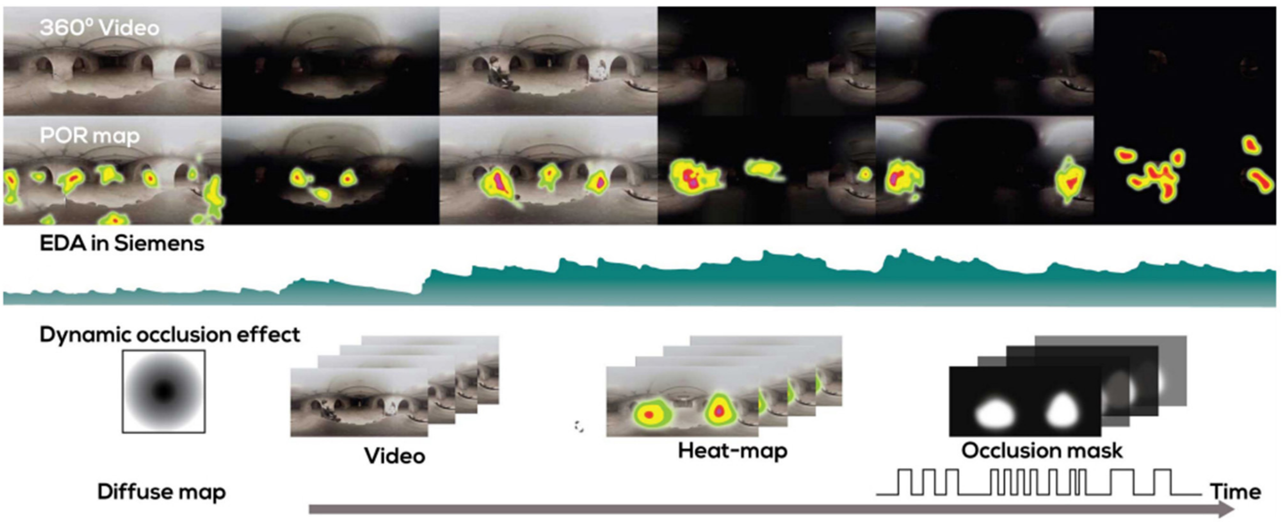


FIGURE 8 | A 360 video scene is manipulated by the user's gaze data and skin conductance response data to increase the users' arousal levels. These arousal levels create darker flashing moments making it more dramatic, similar to those seen in horror movies. To direct the user's focus to the people within the video, the system modulates a surface shader dynamically that occludes locations informed by the gaze tracking system's data as areas that are not of interest to the user.

New Realities

The importance of open-source transparency will likely expand substantially as AIs increasingly become BCI-driven coprocessors for the human mind. Whether algorithms are designed to help homeostatically regulate stress, or to monitor engagement and optimize learning or improve safety, these reality-selection algorithms will likely be critical to examine for integrity, implicit bias, and conflicts of interest. As AIs take on an adjunct homunculus role, their ability to make sense of BCI data and be creatively applied to novel applications will depend on the adaptability, extensibility, and transparency of the BCI tools on which they are built. All this comes with great possibilities to fundamentally transform human consciousness in addition to extreme risks and concerns, and open source may be one of the key factors that helps us navigate this conversation as we build safeguards into this new paradigm.

The issues surrounding algorithm integrity and examples of ways in which open-source code can mitigate those issues can be drawn from the cyber-security sector. For example, WannaCry ransomware targeted a vulnerability in the closed-source Windows operating system (Petrenko et al., 2018) that had existed for over a decade and only came to worldwide attention after the WannaCry crypto-worm infected ∼300,000 computers worldwide, including the U.K.'s National Health Service which cost the organization nearly £100 M in canceled appointments and cleanup. In contrast, the Heartbleed security bug in the open-source OpenSSL cryptography library was discovered and fixed in just over 2 years. While it can be hard to compare any two vulnerabilities, data has suggested that open-source defects are found and fixed more rapidly than closed-source projects (Paulson et al., 2004). The integrity and security of cognition augmenting and reality-selection algorithms is very likely to present a host of cyber-attack opportunities for anything from lone-wolf hackers hawking their wares, to state actors creating individually targeted propaganda. The ability for open-source public review of algorithm integrity may be a way to catch these vulnerabilities in a timely manner.

Implicit bias has been documented in machine learning AI algorithms that govern everything from filtering job applicants to home loan approval (O'neil, 2016; Buolamwini and Gebru, 2018). The implicit bias embodied in the algorithms and machine learning models often ends up reflecting and reinforcing the racial and gender inequities present in our society. As we diversify a multiplicity of virtual and mixed realities, both the risk to exacerbate and the opportunity to mitigate implicit bias will be great. For example, Mel Slater and his students have demonstrated that the embodiment of light-skinned participants in a dark-skinned virtual body significantly reduces implicit racial bias against dark-skinned people, in contrast to embodiment in light-skinned or purple-skinned avatars, or ones with no virtual body at all (Peck et al., 2013). Virtual Reality presents a persuasive tool for potentially placing people into a different body stereotype, particularly race or gender, by modifying the form of their body image. This is accomplished by a setup that is referred to as 'virtual embodiment'. The participants wear a broad field-of-view head-mounted display and when they look down toward themselves in the VR, they see a programmed virtual body (VB) substituting their own real body. They also see this body when looking at their (geometrically accurate) reflection in a virtual mirror. These kinds of virtual reality experiences have the potential to increase empathy and understanding. We speculate based on prior research that as BCIs combined with machine learning become increasingly capable to detect biometric patterns associated with complex cognition such as implicit bias and empathy (Hasler et al., 2017; Levsen and Bartholow, 2018; Luo et al., 2018; Katsumi et al., 2020; Patané et al., 2020), feedback loops between BCIs and virtual reality content will be positioned to either diminish or amplify these internal states. Open-source algorithms and models that can be audited may be an important tool to ensure implicit bias is mitigated and empathy is enhanced as we multiply reality.

Koutsouleris et al. (2020), reported that AI algorithms were able to predict whether people would have a psychotic episode using a combination of clinical, brain imaging, and genetic data. The positive potential of algorithms that could help to intervene or otherwise divert a catastrophic life event for millions of people worldwide cannot be overstated. Looking at an example with less devastating, but potentially broader impact, research has shown that struggles of obesity are driven, in part, by food stimuli that hijack the decision making centers in the brain to create an overwhelming compulsion to eat (Stice et al., 2009; Cobb et al., 2015; Mejova et al., 2015). To short circuit this stimulus-response behavior we can easily imagine how a closed-loop Augmented Reality system that occludes specific food stimuli, like donuts, could be a great aid to those in the process of rebuilding their relationship with food, as shown in **Figure 9**. And yet this technology also presents a clear risk for mistakes and conflicts of interest to have dire consequences. The landscape of BCI-driven reality selection is rife with both utopian and dystopian possibilities. Relapse into addiction, for example, is known to be triggered by specific associative stimuli like a cigarette or a lighter (Shiffman, 2005; Leventhal et al., 2008) and very likely to exhibit greater susceptibility under certain neurological states (Potvin et al., 2015; Witteman et al., 2015). It is easy to imagine a utopian world with reduced substance abuse rates simply by detecting susceptible states and using augmented reality to block trigger stimuli thereby enabling people to be productive members of society. On the contrary, it is equally easy to imagine the dystopian world that might follow if Purdue Pharma [maker of the highly addictive opiate, OxyContin, that drove opiate addiction rates to all-time highs in the United States, (Knisely et al., 2008)] were generating BCI-driven targeted advertising into our reality feeds. We speculate that possible futures of BCI-driven reality selection range from greater safety, health, wellness, rates of learning, and creativity to security risks, manipulated addiction, misinformation, and brainwashing. Transparency may be a critical principle to help unlock the utopian and steer clear of the most dystopian visions of our future.

CLOSING TOPICS

As we have discussed in this paper, BCIs are moving beyond siloed research labs to explore new use-cases ranging from the arts and rehabilitation to gaming and augmented reality. The accessibility, adaptability, and transparency of BCI tools and data will significantly impact how we collectively navigate this new digital age. It is essential that we build technologies that are not only affordable, but that also can be used for

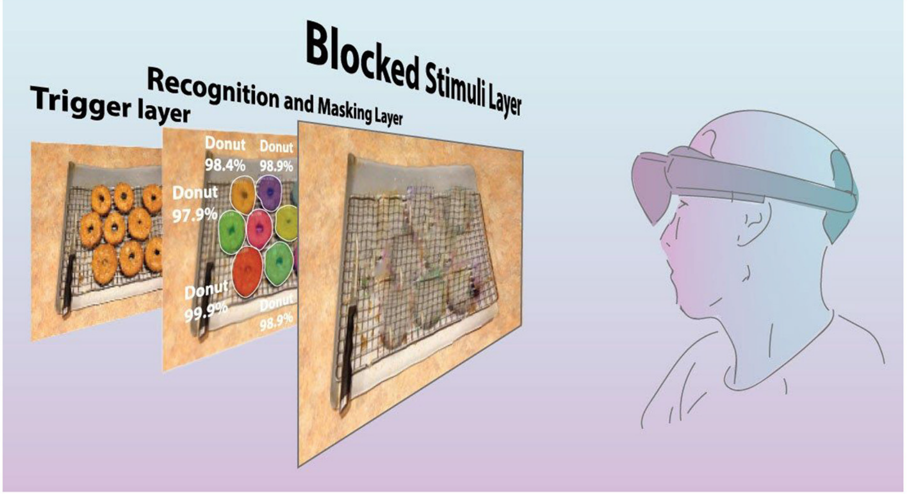


FIGURE 9 | Diagram showing close-loop scenario where stimuli are removed from user perception. The trigger layer shows a stimulus that the user intends to avoid. The recognition layer is computed on the user's device and masks the stimuli to be blocked. The blocked stimuli layer is what the user perceives when a trigger stimulus is shown.

diverse applications while delivering transparency of the data and algorithms employed. Open-source principles would enable BCI technologies to be explored from different perspectives and for novel applications with confidence that the data is relevant and accurately reflecting underlying physiological changes. Enabling BCI technologies to be used in the broadest possible range of applications will ensure that more voices can understand, utilize, and validate the integrity of these signals as we shape a BCI-augmented future.

Both art and technology aim to reshape the world we exist in, re-envisioning what we perceive as real and understanding nature's own limits. For decades, industry has been inspiring new technologies (3D printing, e-paper, satellites) and presenting critical reflections like Black Mirror's "The Entire History of You" that gives us mental frameworks to rethink our relationship with technology. As we move into a future that increasingly merges technology with humanity, the artist's role must be one of an active partner in preparing the direction of research and facilitating synergy between science and technology of science and technology as a vital means for understanding the world.

We speculate that Augmented cognition driven by BCI technology may be poised to transform humanity at a level on par with or exceeding that of the written word. The possibilities to learn faster through dynamic material that is individually tuned to each person's psychophysiology, develop strategies for enhanced creativity and even perhaps bootstrap our biology into new forms of distributed or collective consciousness, may have profound implications for the ability of humanity to understand the universe and its place therein. And yet, much as with any powerful tool, part and parcel with the potential benefits come potential risks. Whether it is the possibility for a future in which we can create reality filters based on physiological responses or read the cataloged memories of alleged criminals (Flisfeder, 2018), there are very real risks as we look forward into a world powered by BCI-driven augmented cognition. Despite these risks, the benefits are too profound and the advantages too immediate to imagine a reverse course. If someone can save themselves or a family member from addiction or a psychotic break (Koutsouleris et al., 2020), or if a driver or pilot can be safer (Healey and Picard, 2005; Zepf et al., 2019), or if a day-trader can be the smartest person in the room, the unrelenting force of progress will likely overpower any attempts to outright halt it. Instead, we may consider building a future based on open-source principles including adaptability, extensibility, and transparency so as to democratize the conversation with different perspectives, deliver the openness needed to conduct replicable science, and understand the algorithms and models that will play increasingly important roles in creating our realities and world views.

Setting up the technical as well as ethical and societal norms of a BCI-driven future will require a diversity of transdisciplinary perspectives. Sitting at the nexus of biology, electrical engineering, and computer science, BCIs are transdisciplinary by their nature, and they also present an opportunity to bring perspectives ranging from psychology, health, and physical education to history, philosophy, and the arts. Bringing together a diversity of ideas and viewpoints can help ensure that this transformative technology is set up to serve not only the most privileged members of society, but also enable individual ingenuity to go beyond preconceived use-cases to solve issues that transcend physical, economic, and cultural boundaries. Even more than the sum of siloed individual perspectives, creating transdisciplinary conversations that explore the intersections between different points of view can multiply the information and ideas to imagine our future realities (Nijholt et al., 2018; Dikker et al., 2019b). Those conversations might become more common and start at a younger age, by bringing BCIs into K-12 project-based curricula and into hackathons that promote diverse teams including artists and philosophers as well as engineers and scientists. By building BCI tools that are adaptable and transparent as well as accessible, the resulting applications and conversations can go beyond preconceived use-cases to explore the widest scope of possibilities that BCIs may unlock for the future. Greater openness may require some reframing of the solution space in the context of principles including those of adaptability, extensibility and transparency. Developing BCI tools that are adaptable and extensible can democratize the development of new ideas and applications to imagine beyond the board-room developed use-cases. Cultivating more transparency, with greater access to raw data, source code, and visibility into black-box models can facilitate creating replicable scientific knowledge, and a trust that future realities will be secure and serve the interests of all.

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/s.

AUTHOR CONTRIBUTIONS

GB and SM conceived of the presented idea, drafted the manuscript, and designed the figures. All authors discussed the results and commented on the manuscript.

REFERENCES

Abramovic, M., Dikker, S., and Oosterik, M. (2014). Measuring the magic of mutual gaze.

Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., and Adeli, H. (2018). Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Comput. Biol. Med.* 100, 270–278. doi: 10.1016/j.compbiomed.2017.09.017

Almogbel, M. A., Dang, A. H., and Kameyama, W. (2018). "EEG-signals based cognitive workload detection of vehicle driver using deep learning," in *2018 20th International Conference on Advanced Communication Technology (ICACT) (IEEE)* (Chuncheon), 256–259. doi: 10.23919/ICACT.2018.8323716

Appelhans, B. M., and Luecken, L. J. (2006). Heart rate variability as an index of regulated emotional responding. *Rev. Gen. Psychol.* 10, 229–240. doi: 10.1037/1089-2680.10.3.229

Baltatzis, V., Bintsi, K.-M., Apostolidis, G. K., and Hadjileontiadis, L. J. (2017). Bullying incidences identification within an immersive environment using HD EEG-based analysis: a swarm decomposition and deep learning approach. *Sci. Rep.* 7, 1–8. doi: 10.1038/s41598-017-17562-0

Banzi, M., Cuartielles, D., Igoe, T., Martino, G., and Mellis, D. (2015). *Arduino: Open-Source Project Official Website*. Arduino.

Behrman, D. (1998). Wave Train.

Bernal, G., and Maes, P. (2017). "Emotional beasts: visually expressing emotions through avatars in VR," in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems CHI EA'17* (New York, NY: ACM), 2395–2402. doi: 10.1145/3027063.3053207

Bernal, G., Yang, T., Jain, A., and Maes, P. (2018). "PhysioHMD: a conformable, modular toolkit for collecting physiological data from head-mounted displays," in *Proceedings of the 2018 ACM International Symposium on Wearable Computers ISWC'18* (New York, NY: Association for Computing Machinery), 160–167. doi: 10.1145/3267242.3267268

Berntson, G. G., Bigger, J. T., Eckberg, D. L., Grossman, P., Kaufmann, P. G., Malik, M., et al. (1997). Heart rate variability: origins, methods, and interpretive caveats. *Psychophysiology* 34, 623–648. doi: 10.1111/j.1469-8986.1997.tb02140.x

Biswas, D., Simões-Capela, N., Hoof, C. V., and Helleputte, N. V. (2019). Heart rate estimation from wrist-worn photoplethysmography: a review. *IEEE Sens. J.* 19, 6560–6570. doi: 10.1109/JSEN.2019.2914166

Buolamwini, J., and Gebru, T. (2018). "Gender shades: intersectional accuracy disparities in commercial gender classification," in *Conference on Fairness, Accountability and Transparency (PMLR)* (New York NY), 77–91. Available online at: http://proceedings.mlr.press/v81/buolamwini18a.html (accessed January 22, 2021).

Burke, J. F., Merkow, M. B., Jacobs, J., Kahana, M. J., and Zaghloul, K. A. (2015). Brain computer interface to enhance episodic memory in human participants. *Front. Hum. Neurosci.* 8:1055. doi: 10.3389/fnhum.2014.01055

Buzsaki, G. (2006). *Rhythms of the Brain*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780195301069.001.0001

Chinese Primary School Halts Trial of Device That Monitors Pupils, Brainwaves (2019). *The Guardian*. Available online at: http://www.theguardian.com/world/2019/nov/01/chinese-primary-school-halts-trial-of-device-that-monitors-pupils-brainwaves (accessed January 22, 2021).

Cobb, L. K., Appel, L. J., Franco, M., Jones-Smith, J. C., Nur, A., and Anderson, C. A. (2015). The relationship of the local food environment with obesity: a systematic review of methods, study quality, and results. *Obesity* 23, 1331–1344. doi: 10.1002/oby.21118

Dan, C. (2011). *ZeoScope*. Available online at: https://github.com/dancodru/ZeoScope (accessed January 17, 2021).

Dikker, S., Michalareas, G., Oostrik, M., Serafimaki, A., Kahraman, H. M., Struiksma, M. E., et al. (2019a). Crowdsourcing neuroscience: inter-brain coupling during face-to-face interactions outside the laboratory. *NeuroImage* 227:117436. doi: 10.1016/j.neuroimage.2020.117436

Dikker, S., Montgomery, S., and Tunca, S. (2019b). "Using synchrony-based neurofeedback in search of human connectedness," in *Brain Art*, eds A. Nijholt (Cham: Springer), 161–206. doi: 10.1007/978-3-030-14323-7_6

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., et al. (2017). Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Curr. Biol.* 27, 1375–1380. doi: 10.1016/j.cub.2017.04.002

Electrophysiology, T. F. (1996). Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *Circulation* 93, 1043–1065. doi: 10.1161/01.CIR.93.5.1043

Engel, A. K., Fries, P., and Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top–down processing. *Nat. Rev. Neurosci.* 2, 704–716. doi: 10.1038/35094565

Faivre, N., Salomon, R., and Blanke, O. (2015). Visual consciousness and bodily self-consciousness. *Curr. Opin. Neurol.* 28, 23–28. doi: 10.1097/WCO.0000000000000160

Flisfeder, M. (2018). Black Mirror, "playtest," and the crises of the present. *Black Mirror Crit. Media Theory* 141:141. doi: 10.1177/1461444810365313

Fries, P. (2015). Rhythms for cognition: communication through coherence. *Neuron* 88, 220–235. doi: 10.1016/j.neuron.2015.09.034

Gray, C. M., König, P., Engel, A. K., and Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* 338, 334–337. doi: 10.1038/338334a0

Griffith, A. (2006). *An Exploration of the OpenEEG Project*. CHG Wright's BioData Syst.

Gupta, K., Hajika, R., Pai, Y. S., Duenser, A., Lochner, M., and Billinghurst, M. (2020). "Measuring human trust in a virtual assistant using physiological sensing in virtual reality," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Atlanta, GA), 756–765. doi: 10.1109/VR46266.2020.00099

Hasler, B. S., Spanlang, B., and Slater, M. (2017). Virtual race transformation reverses racial in-group bias. *PLoS ONE* 12:e0174965. doi: 10.1371/journal.pone.0174965

Healey, J. A., and Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Trans. Intell. Transp. Syst.* 6, 156–166. doi: 10.1109/TITS.2005.848368

Hoffman, C., and Bast, G. (2017). *Cosa Mentale, Draw With Two Brains Connected to One Hand | Mentalista*. Available online at: https://mentalista.fr/en/cosamentale (accessed October 31, 2020).

Höller, Y., Uhl, A., Bathke, A., Thomschewski, A., Butz, K., Nardone, R., et al. (2017). Reliability of EEG measures of interaction: a paradigm shift is needed to fight the reproducibility crisis. *Front. Hum. Neurosci.* 11:441. doi: 10.3389/fnhum.2017.00441

How to Hack Toy EEGs | Frontier Nerds (2010). Available online at: http://www.frontiernerds.com/brain-hack (accessed January 17, 2021).

Huang, D., Wang, X., Li, J., and Tang, W. (2020). "Virtual reality for training and fitness assessments for construction safety," in *2020 International Conference on Cyberworlds (CW)* (Caen), 172–179. doi: 10.1109/CW49994.2020.00036

Kaplan, R., Bush, D., Bonnefond, M., Bandettini, P. A., Barnes, G. R., Doeller, C. F., et al. (2014). Medial prefrontal theta phase coupling during spatial memory retrieval. *Hippocampus* 24, 656–665. doi: 10.1002/hipo.22255

Katsumi, Y., Dolcos, F., Moore, M., Bartholow, B. D., Fabiani, M., and Dolcos, S. (2020). Electrophysiological correlates of racial in-group bias in observing nonverbal social encounters. *J. Cogn. Neurosci.* 32, 167–186. doi: 10.1162/jocn_a_01475

Kim, B., Khanna, R., and Koyejo, O. O. (2016). "Examples are not enough, learn to criticize! criticism for interpretability," in *Advances in Neural Information Processing Systems*, ed N. Thakor (Springer), 2280–2288.

Knisely, J. S., Wunsch, M. J., Cropsey, K. L., and Campbell, E. D. (2008). Prescription opioid misuse index: a brief questionnaire to assess misuse. *J. Subst. Abuse Treat.* 35, 380–386. doi: 10.1016/j.jsat.2008.02.001

Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* 17, 307–321. doi: 10.1038/nrn.2016.22

Kos, M., Li, X., Khaghani-Far, I., Gordon, C. M., Pavel, M., and Jimison Member, H. B. (2017). Can accelerometry data improve estimates of heart rate variability from wrist pulse PPG sensors? *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.* 2017, 1587–1590. doi: 10.1109/EMBC.2017.8037141

Koutsouleris, N., Dwyer, D. B., Degenhardt, F., Maj, C., Urquijo-Castro, M. F., Sanfelici, R., et al. (2020). Multimodal machine learning workflows for prediction of psychosis in patients with clinical high-risk syndromes and recent-onset depression. *JAMA Psychiatry* 78, 195–209. doi: 10.1001/jamapsychiatry.2020.3604

Kropotov, J. D. (2010). *Quantitative EEG, Event-Related Potentials and Neurotherapy*. Academic Press.

Lécuyer, A., Lotte, F., Reilly, R. B., Leeb, R., Hirose, M., and Slater, M. (2008). Brain-computer interfaces, virtual reality, and videogames. *Computer* 41, 66–72. doi: 10.1109/MC.2008.410

Leslie, G. (2020). Inner rhythms: vessels as a sustained brain-body performance practice. *Leonardo*, 1–8. doi: 10.1162/leon_a_01963

Leventhal, A. M., Waters, A. J., Breitmeyer, B. G., Miller, E. K., Tapia, E., and Li, Y. (2008). Subliminal processing of smoking-related and affective stimuli in tobacco addiction. *Exp. Clin. Psychopharmacol.* 16, 301–312. doi: 10.1037/a0012640

Levsen, M. P., and Bartholow, B. D. (2018). Neural and behavioral effects of regulating emotional responses to errors during an implicit racial bias task. *Cogn. Affect. Behav. Neurosci.* 18, 1283–1297. doi: 10.3758/s13415-018-0639-8

Llinás, R., Ribary, U., Contreras, D., and Pedroarena, C. (1998). The neuronal basis for consciousness. *Philos. Trans. R. Soc. B Biol. Sci.* 353, 1841–1849. doi: 10.1098/rstb.1998.0336

Lorenzetti, V., Melo, B., Basílio, R., Suo, C., Yücel, M., Tierra-Criollo, C. J., et al. (2018). Emotion regulation using virtual environments and real-time fMRI neurofeedback. *Front. Neurol.* 9:390. doi: 10.3389/fneur.2018.00390

Lozano-Hemmer, R. (2008). Pulse Spiral.

Lundberg, S. M., and Lee, S.-I. (2017). A Unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* 30, 4765–4774.

Luo, S., Han, X., Du, N., and Han, S. (2018). Physical coldness enhances racial in-group bias in empathy: electrophysiological evidence. *Neuropsychologia* 116, 117–125. doi: 10.1016/j.neuropsychologia.2017.05.002

Marshall, D., Coyle, D., Wilson, S., and Callaghan, M. (2013). Games, gameplay, and BCI: the state of the art. *IEEE Trans. Comput. Intell. AI Games* 5, 82–99. doi: 10.1109/TCIAIG.2013.2263555

Masui, K., Nagasawa, T., and Tsumura, N. (2020). "Continuous estimation of emotional change using multimodal affective responses," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 290–291. doi: 10.1109/CVPRW50498.2020.00153

Max, S. (1977). Public Supply.

McNutt, M. (2014). Reproducibility. *Science* 343, 229–229. doi: 10.1126/science.1250475

Mejova, Y., Haddadi, H., Noulas, A., and Weber, I. (2015). "Foodporn: obesity patterns in culinary interactions," in *Proceedings of the 5th International Conference on Digital Health* (New York, NY), 51–58. doi: 10.1145/2750511.2750524

Meyer, L., Grigutsch, M., Schmuck, N., Gaston, P., and Friederici, A. D. (2015). Frontal-posterior theta oscillations reflect memory retrieval during sentence comprehension. *Cortex J. Devoted Study Nerv. Syst. Behav.* 71, 205–218. doi: 10.1016/j.cortex.2015.06.027

Miltner, W. H., Braun, C., Arnold, M., Witte, H., and Taub, E. (1999). Coherence of gamma-band EEG activity as a basis for associative learning. *Nature* 397, 434–436. doi: 10.1038/17126

Montana, J. I., Matamala-Gomez, M., Maisto, M., Mavrodiev, P. A., Cavalera, C. M., Diana, B., et al. (2020). The benefits of emotion regulation interventions in virtual reality for the improvement of wellbeing in adults and older adults: a systematic review. *J. Clin. Med.* 9:500. doi: 10.3390/jcm9020500

Montgomery, S. (2018a). *Hive Mind, Don't Speak | Neural. Winter*. Available online at: http://neural.it/2018/05/hive-mind-dont-speak/ (accessed November 13, 2020).

Montgomery, S. (2018b). *Synergy in Art and Science Could Save the World*. San Francisco, CA. Available online at: https://www.youtube.com/watch?v=OXFxVWI6QNU (accessed January 17, 2021).

Montgomery, S., and Laefsky, I. (2011). Bio-Sensing: hacking the doors of perception. *Make Mag.* 26, 104–111.

Montgomery, S., Rioja, D., and Bagdatli, M. (2011). "Emergence," in *ISEA2011: 17th International Symposium on Electronic Art (Istambul)*. Available online at: https://isea-archives.siggraph.org/art-events/sean-montgomery-diego-rioja-mustafa-bagdatli-emergence/ (accessed January 17, 2021).

Montgomery, S. M. (2010). "Measuring biological signals: concepts and practice," in *Proceedings of the Fourth International Conference on Tangible, Embedded, and Embodied Interaction Tei'10* (New York, NY: Association for Computing Machinery), 337–340. doi: 10.1145/1709886.1709969

Morelli, D., Bartoloni, L., Colombo, M., Plans, D., and Clifton, D. A. (2018). Profiling the propagation of error from PPG to HRV features in a wearable physiological-monitoring device. *Healthc. Technol. Lett.* 5, 59–64. doi: 10.1049/htl.2017.0039

Murphy, J., and Russomanno, C. (2013). *OpenBCI. Wikipedia*. Available online at: https://openbci.com/ (accessed October 31, 2020).

Nijboer, F., Allison, B. Z., Dunne, S., Bos, D. P.-O., Nijholt, A., and Haselager, P. (2011). "A preliminary survey on the perception of marketability of brain-computer interfaces and initial development of a repository of BCI companies," in *Proceedings 5th International Brain-Computer Interface Conference (BCI 2011)* (Verlag der Technischen Universität Graz), 344–347.

Nijholt, A., Bos, D. P.-O., and Reuderink, B. (2009). Turning shortcomings into challenges: brain–computer interfaces for games. *Entertain. Comput.* 1, 85–94. doi: 10.1016/j.entcom.2009.09.007

Nijholt, A., Jacob, R. J., Andujar, M., Yuksel, B. F., and Leslie, G. (2018). "Brain-computer interfaces for artistic expression," in *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal, QC), 1–7. doi: 10.1145/3170427.3170618

O'neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown.

Ossmann, R., Parker, S., Thaller, D., Pecyna, K., García-Soler, A., Morales, B., et al. (2014). AsTeRICS, a flexible AT construction set. *Int. J. Adapt. Control Signal Process.* 28, 1475–1503. doi: 10.1002/acs.2496

Patané, I., Lelgouarch, A., Banakou, D., Verdelet, G., Desoche, C., Koun, E., et al. (2020). Exploring the effect of cooperation in reducing implicit racial bias and its relationship with dispositional empathy and political attitudes. *Front. Psychol.* 11:510787. doi: 10.3389/fpsyg.2020.510787

Paulson, J. W., Succi, G., and Eberlein, A. (2004). An empirical study of open-source and closed-source software products. *IEEE Trans. Softw. Eng.* 30, 246–256. doi: 10.1109/TSE.2004.1274044

Pearlman, E. L. (2016). Noor-A Brain Opera. in Microwave Festival.

Peck, T. C., Seinfeld, S., Aglioti, S. M., and Slater, M. (2013). Putting yourself in the skin of a black avatar reduces implicit racial bias. *Conscious. Cogn.* 22, 779–787. doi: 10.1016/j.concog.2013.04.016

Petrenko, A. S., Petrenko, S. A., Makoveichuk, K. A., and Chetyrbok, P. V. (2018). Protection model of PCS of subway from attacks type ≪wanna cry≫,≪petya≫ and ≪bad rabbit≫ IoT. in *2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)* (IEEE), 945–949.

Picard, R. W. (1995). *Affective Computing-MIT Media Laboratory Perceptual Computing Section Technical Report No. 321*. Camb.

Potvin, S., Tikàsz, A., Dinh-Williams, L. L.-A., Bourque, J., and Mendrek, A. (2015). Cigarette cravings, impulsivity, and the brain. *Front. Psychiatry* 6:125. doi: 10.3389/fpsyt.2015.00125

Primary School in China Suspends Use of BrainCo Brainwave Tracking Headband (2019). *Glob. Shak*. Available online at: https://globalshakers.com/primary-school-in-china-suspends-use-of-brainco-brainwave-tracking-headband/ (accessed January 22, 2021).

Produce Consume Robot, and LoVid (2013). *Telephone Rewired at Pocket Gallery*. Available online at: https://www.harvestworks.org/telephone-rewired/ (accessed January 17, 2021).

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). *"Why Should I Trust You?": Explaining the Predictions of Any Classifier*. Available online at: http://arxiv.org/abs/1602.04938 (accessed January 21, 2021).

Sandberg, K., Timmermans, B., Overgaard, M., and Cleeremans, A. (2010). Measuring consciousness: is one measure better than the other? *Conscious. Cogn.* 19, 1069–1078. doi: 10.1016/j.concog.2009.12.013

Schneider, R. (2010). *Mindflex Hack Turns Brain Waves Into Music. Wired*. Available online at: https://www.wired.com/2010/10/robert-schneider-teletron/ (accessed January 21, 2021).

Shcherbina, A., Mattsson, C. M., Waggott, D., Salisbury, H., Christle, J. W., Hastie, T., et al. (2017). Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. *J. Pers. Med.* 7:3. doi: 10.3390/jpm7020003

Shiffman, S. (2005). Dynamic influences on smoking relapse process. *J. Pers.* 73, 1715–1748. doi: 10.1111/j.0022-3506.2005.00364.x

Sitsky, L. (2002). *Music of the Twentieth-Century Avant-Garde: A Biocritical Sourcebook: A Biocritical Sourcebook*. ABC-CLIO.

Sourina, O., Liu, Y., Wang, Q., and Nguyen, M. K. (2011). "EEG-based personalized digital experience," in *International Conference on Universal Access in Human-Computer Interaction* (Orlando, FL: Springer), 591–599. doi: 10.1007/978-3-642-21663-3_64

Stice, E., Spoor, S., Ng, J., and Zald, D. H. (2009). Relation of obesity to consummatory and anticipatory food reward. *Physiol. Behav.* 97, 551–560. doi: 10.1016/j.physbeh.2009.03.020

Sung, Y., Cho, K., and Um, K. (2012). A development architecture for serious games using BCI (brain computer interface) sensors. *Sensors* 12, 15671–15688. doi: 10.3390/s121115671

Uhlhaas, P., Pipa, G., Lima, B., Melloni, L., Neuenschwander, S., Nikoli,ć, D., et al. (2009). Neural synchrony in cortical networks: history, concept and current status. *Front. Integr. Neurosci.* 3.:17 doi: 10.3389/neuro.07.017.2009

View Research, G. (2019). *Brain Computer Interface Market Size Report, 2020–2027*. Available online at: https://www.grandviewresearch.com/industry-analysis/brain-computer-interfaces-market (accessed January 22, 2021).

Vujic, A., Tong, S., Picard, R., and Maes, P. (2020). "Going with our guts: potentials of wearable electrogastrography (EGG) for affect detection," in *Proceedings of the 2020 International Conference on Multimodal Interaction* (Netherlands), 260–268. doi: 10.1145/3382507.3418882

Wall Street Journal (2019). *How China Is Using Artificial Intelligence in Classrooms | WSJ*. Available online at: https://www.youtube.com/watch?v=JMLsHI8aV0g (accessed January 22, 2021).

Wang, J., Wang, W., and Hou, Z.-G. (2019). Toward improving engagement in neural rehabilitation: attention enhancement based on brain–computer interface and audiovisual feedback. *IEEE Trans. Cogn. Dev. Syst.* 12, 787–796. doi: 10.1109/TCDS.2019.2959055

Witteman, J., Post, H., Tarvainen, M., de Bruijn, A., Perna, E. D. S. F., Ramaekers, J. G., et al. (2015). Cue reactivity and its relation to craving and relapse in alcohol dependence: a combined laboratory and field study. *Psychopharmacology* 232, 3685–3696. doi: 10.1007/s00213-015-4027-6

Wolpaw, J. R., Birbaumer, N., Heetderks, W. J., McFarland, D. J., Peckham, P. H., Schalk, G., et al. (2000). Brain-computer interface technology: a review of the first international meeting. *IEEE Trans. Rehabil. Eng.* 8, 164–173. doi: 10.1109/TRE.2000.847807

WorldFamousElectronicst (2011). *Pulse Sensor*. Available online at: https://github.com/WorldFamousElectronics/PulseSensorStarterProject (accessed January 23, 2021).

Zepf, S., Dittrich, M., Hernandez, J., and Schmitt, A. (2019). "Towards empathetic car interfaces: emotional triggers while driving," in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow; Scotland), 1–6. doi: 10.1145/3290607.3312883

Conflict of Interest: SM was employed by company Connected Future Labs.

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.

MINI REVIEW
published: 30 April 2021
doi: 10.3389/fcomp.2021.661178

EEG-Based Auditory Attention Detection and Its Possible Future Applications for Passive BCI

Joan Belo[1]*, *Maureen Clerc*[1] *and Daniele Schön*[2]

[1] *Inria, Université Côte d'Azur, Valbonne, France,* [2] *Dynamics of Cognitive Processes Group (DCP), CNRS, INS, Aix-Marseille Université, Marseille, France*

The ability to discriminate and attend one specific sound source in a complex auditory environment is a fundamental skill for efficient communication. Indeed, it allows us to follow a family conversation or discuss with a friend in a bar. This ability is challenged in hearing-impaired individuals and more precisely in those with a cochlear implant (CI). Indeed, due to the limited spectral resolution of the implant, auditory perception remains quite poor in a noisy environment or in presence of simultaneous auditory sources. Recent methodological advances allow now to detect, on the basis of neural signals, which auditory stream within a set of multiple concurrent streams an individual is attending to. This approach, called EEG-based auditory attention detection (AAD), is based on fundamental research findings demonstrating that, in a multi speech scenario, cortical tracking of the envelope of the attended speech is enhanced compared to the unattended speech. Following these findings, other studies showed that it is possible to use EEG/MEG (Electroencephalography/Magnetoencephalography) to explore auditory attention during speech listening in a Cocktail-party-like scenario. Overall, these findings make it possible to conceive next-generation hearing aids combining customary technology and AAD. Importantly, AAD has also a great potential in the context of passive BCI, in the educational context as well as in the context of interactive music performances. In this mini review, we firstly present the different approaches of AAD and the main limitations of the global concept. We then expose its potential applications in the world of non-clinical passive BCI.

Keywords: AAD, EEG, auditory attention, passive BCI, education, art

Edited by:
Camille Jeunet, UMR5287 Institut de Neurosciences Cognitives et Intégratives d'Aquitaine (INCIA), France

Reviewed by:
Lucia Pepa, Marche Polytechnic University, Italy
Jan Mucha, Brno University of Technology, Czechia

***Correspondence:**
Joan Belo
joan.belo@inria.fr

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Computer Science

Received: *30 January 2021*
Accepted: *06 April 2021*
Published: *30 April 2021*

Citation:
Belo J, Clerc M and Schön D (2021) EEG-Based Auditory Attention Detection and Its Possible Future Applications for Passive BCI. Front. Comput. Sci. 3:661178. doi: 10.3389/fcomp.2021.661178

INTRODUCTION

The ability to discriminate and attend one specific sound source in a complex auditory environment is of utmost importance in the animal world both in terms of avoiding dangers and finding mates. In humans, this ability goes well-beyond survival and reproduction since it is a fundamental skill for efficient communication. Indeed, it allows us to follow a family conversation or discuss with a friend in a bar. In music, this ability is challenged by the simultaneous layering of several instruments playing together, requiring sound source segregation to fully appreciate the ensemble. This ability is also challenged in hearing-impaired individuals and more precisely in those with a cochlear implant (CI). Indeed, due to the limited spectral resolution of the implant, auditory perception remains quite poor in a noisy environment or in presence of simultaneous auditory sources. Thus, being able to enhance the relevant/attended source would facilitate source separation in individuals with CI. However, monitoring the attended auditory source is not easy, as this changes in time.

Recent methodological advances allow now to detect, on the basis of neural signals, which auditory stream within a set of multiple concurrent streams an individual is attending to. This approach, called EEG-based auditory attention detection (AAD), is based on fundamental research findings demonstrating that, in a multi speech scenario, cortical tracking of the envelope of the attended speech is enhanced compared to the unattended speech (Mesgarani et al., 2009; Ding and Simon, 2012; Mesgarani and Chang, 2012; Pasley et al., 2012; Zion Golumbic et al., 2013). Following these findings, other studies showed that it is possible to use EEG/MEG to explore auditory attention during speech listening in a Cocktail-party-like scenario (Ding and Simon, 2012; O'Sullivan et al., 2015; Akram et al., 2016). This field of research has grown rapidly and several new methods and techniques were developed in the last years to improve the first attempts.

Overall, these findings make it possible to conceive next-generation hearing aids combining customary technology and AAD. Importantly, AAD has also a great potential in the context of passive BCI, in the educational context as well as in the context of interactive music performances.

In this mini review, we firstly present the different approaches of AAD and the main limitations of the global concept. We then expose its potential applications in the world of non-clinical passive BCI.

The main rationale behind this mini-review is to bridge the EEG-based AAD and Passive BCI communities and to provide insights about how the emerging synergy will develop. While previous reviews have been published on technical aspects of AAD, this mini-review attempts to briefly present EEG-based AAD in a broader perspective and to guide the reader to the most relevant sources. The methodology used to find and include papers in the current mini-review was as follows. The search was carried on using both Pubmed and Google Scholar. Keywords included machine learning, decoding, encoding, auditory attention, EEG, and speech. Pubmed gave 88 results and Scholar 8,460 results. These results were then filtered with the following exclusion criteria: articles about engineering techniques that are not directly in relation with EEG-based AAD methodology, articles with methods that were not applied to M/EEG data, articles that were not published in a peer-review journal, articles that were cited <1 time. This reduced the number of included articles to 20 (see **Table 1**).

EEG-BASED AUDITORY ATTENTION DETECTION METHODS

There are many different AAD methods based on EEG measures. Identifying the attended speaker using cortical activity measurement is possible because the amplitude envelope of the speech stream (a crucial feature for speech comprehension) is represented in the theta and gamma oscillatory activity in the human auditory cortex (Nourski et al., 2009; Giraud and Poeppel, 2012; Kubanek et al., 2013). Attending a source thus results in greater coupling between the envelope of the source and the envelope of neural activity in these bands.

The vast majority of the studies that explored EEG-based AAD performances used two concurrent spatially separated talkers but some of them have explored the impact of speaker number and their location in auditory scene (Schäfer et al., 2018), background noise (Das et al., 2018), reverberation (Fuglsang et al., 2017), number of EEG electrodes (Mirkovic et al., 2015; Bleichner et al., 2016), or even their location (Fiedler et al., 2017) on the performance of AAD algorithms.

One can distinguish two main categories of approaches to detect auditory attention: linear and non-linear models (see Geirnaert et al., 2020 for a comprehensive review of AAD Algorithms).

Linear Models

In the community of linear models, two main "philosophies" are in competition (see Alickovic et al., 2019 for a complete review on linear models): forward, or encoding (encoding because these models are a description of how the system *encodes* information), and backward, or decoding, models.

The objective of the forward strategy is to predict the neural response in the neural data (i.e., EEG channels) from the representation of the audio signal via a temporal response function (i.e., an encoder) that describes the linear relationship between a set of neural data and an audio stimulus at certain time points (Crosse et al., 2016). In the simplest case (i.e., one audio signal) a unique representation of the audio signal is created. This representation can be the amplitude envelope (O'Sullivan et al., 2015), the spectrogram of speech signal (O'Sullivan et al., 2017), or the Mel spectrogram for a music signal (Cantisani et al., 2019). Depending on the type of the chosen representation the analysis can be either univariate (an amplitude envelope is a univariate stimulus feature) or multivariate (a spectrogram is a multivariate stimulus feature). Although it is possible to use multivariate TRF with the forward approach, this strategy is, by nature, univariate (Crosse et al., 2016). Afterward, the audio representation is convolved with an unknown channel-specific TRF. To estimate the TRF (i.e., fit the model parameters), an error minimization is performed between the neural response and the one predicted by the convolution (e.g., Mean-Squared Error) using assumptions about noise distribution (Holdgraf et al., 2017). Once the model's parameters have been estimated, the model is validated on new data. These new data could be from the same dataset used to estimate the parameters (leave-n-out procedure) or from data recorded separately. The validation step is crucial because, to be interpretable, the model should be compatible with new data and make accurate predictions (generalization ability). Finally, the rationale of the forward strategy, in auditory research, is to predict neural data on the basis of the sound's features.

Backward models work similarly but by predicting the auditory representation based on neural data (Alickovic et al., 2019). A pre-trained neural linear decoder is applied to the neural data to reconstruct the chosen representation (this is the reason why this type of approach is sometimes called "*stimulus reconstruction*"). The reconstructed representation is compared to the original representations. A high similarity (correlation) indicates a good performance of the model. Two other approaches can also be mentioned: Canonical Correlation

TABLE 1 | Table describing main important characteristics of AAD reviewed articles.

Article	Data	Method	Subject	Audio features	AV model goodness	AV classification accuracy	Decision window
Akram et al., 2016	MEG	Non-linear (SSM)	11	Amp Env	–	74% (Not sure)	60 s (Not sure)
Bleichner et al., 2016	EEG (+ cEEGrid)	ERP classification	20	–	–	70% (EEG)–66% (cEEGrid)	–
Cantisani et al., 2019	EEG	Linear (SR)	8	Amp Env (AE), Magnitude Spec (MAG), Mel Spec (MEL)	$r = 0.054$ (AE), $r = 0.215$ (MAG), $r = 0.119$ (MEL)	F1 score = 51 (AE), 72 (MAG), 73 (MEL)	24 s
Ciccarelli et al., 2019	EEG	Linear (SR) and Non-linear (DNN)	11	Amp Env	–	66% (Linear), 81% (Non-linear)	10 s
Das et al., 2018	EEG	Linear (SR)	28	Amp Env	$r = \sim 0.06$ (Speaker separation = 10°, SNR = −7.1dB)–$r = \sim 0.14$ (Speaker separation = 180°, SNR = −1.1 dB) [Attended speaker]	97% (Speaker separation = 180°, SNR = −1.1 dB)–59% (Speaker separation = 10°, SNR = −7.1 dB)	30 s
de Cheveigné et al., 2018	EEG	Linear (CCA)	8	Amp Env	$r = \sim 0.3$	~95–~75% (Best CC pairs)	60–10 s
de Taillez et al., 2017	EEG	Non-linear (NN)	20	Amp Env	–	97.6–67.8%	60–2 s
Vandecapelle et al., 2020	EEG	Non-linear (CNN:D, CNN:S+D, CCN:S)	16	Amp Env	–	87% (CNN:S+D), 78% (CNN:D), 70.5% (CNN:S) [subject specific]	10 s
Ding and Simon, 2012	MEG	Linear (SR)	20	Amp Env	$r = \sim 0.2$	–	–
Fiedler et al., 2017	EEG (+ in-Ear-EEG)	Linear (forward)	7	Amp Env	$r = 0.04$	70%	60 s
Fuglsang et al., 2017	EEG	Linear (SR)	26	Amp Env	$r = \sim 0.07$	87.1%	40–50 s
Mesgarani and Chang, 2012	ECoG	Linear (SR)	3	Amp Env	$r = \sim 0.60$	93.0%	NC
Miran et al., 2018a,b	EEG and MEG	Linear (SSM)	3 (EEG)–9 (MEG)	Amp Env	–	70% (MEG data), 80% (EEG data)	1.5 s
Mirkovic et al., 2015	EEG	Linear (SR)	12	Amp Env	–	88.02%	–
O'Sullivan et al., 2015	EEG	Linear (SR)	40	Amp Env	$r = 0.054$ (Subject-specific decoder)	89% [Subject-specific]	60 s
O'Sullivan et al., 2017	ECoG	Non-linear (DNN)	6	Spec	$r = \sim 0.4$ (Attended speaker)	>70% (3 Subjects)	15 s
Pasley et al., 2012	ECoG	Linear (SR) and Non-linear ()	15	Spec	$r = 0.2$–0.3	–	–
Schäfer et al., 2018	EEG	Linear (SR)	10	Amp Env	–	61.1%	30 s
Vandecapelle et al., 2020	EEG	Non-linear (CNN)	16	Pre-processed EEG signal	–	85.1–80.8% [Subject specific]	10–1 s
Zion Golumbic et al., 2013	ECoG	Linear (SR)	6	Amp Env	$r = \sim 0.15$	–	–

Articles are sorted by alphabetical order. The method row indicates the type of model used in the article. Amp Env, Amplitude Envelope; Spec, spectrogram; AV, Average; SR, Stimulus Reconstruction; SSM, State-Space Model; DNN, Deep Neural Network; CNN, Convolutional Neural Network; CCA, Canonical Correlation Analysis; s, second.

Analysis (CCA) and Bayesian state-space modeling. Canonical Correlation Analysis is a hybrid model that combines a decoding and an encoding model. This approach, developed by de Cheveigné et al. (2018), aims to minimize the irrelevant variance in both neural data and stimulus by a linear transformation. Concerning Bayesian state-space modeling (Miran et al., 2018b), it is composed of three modules: a *dynamic encoder/decoder estimation* module, an *attention marker extraction* module, and a *real-time state-space estimator* module (see Miran et al., 2018a for a complete description of the model) and this approach was developed in the purpose of real-time decoding of auditory attention.

As mentioned before, in the context of AAD, linear models are generally used with two (or more) concurrent speech streams in order to determine which stream the listener is attending to. In this case, a representation of each auditory source is created (e.g., speaker 1 and speaker 2). Once the model has been fitted, no matter which approach was chosen, a two-class classifier is used to decide which of the two streams the participant was focused on. To do so, the classifier compares the correlation coefficients between the model output and the original model input representations (e.g., the correlation between the reconstructed envelope and the original audio signals envelopes in backward strategy) over a certain portion of data (decision time windows). The highest correlation indicates which stream the participant was attending to. The length of the decision time window is a crucial parameter because correlation-based measures need a certain amount of information to perform well. However, short decision time windows (<2 s of data) are of interest in BCI for real-time classification.

Generally, AAD performances are assessed with two accuracy metrics: regression accuracy and classification accuracy (Wong et al., 2018). Regression accuracy evaluates the goodness of fit of the model and it is expressed in terms of correlation coefficient (Pearson's correlation, often ranging 0.1–0.2) between the output of the model and the real value (e.g., speech envelope is correlated with reconstructed envelope for backward models). Classification accuracy, on the other hand, evaluates the ability of the classifier to correctly identify the attended stream for a given decision time window and it is generally expressed in terms of percentage of good classification. Classification accuracy is generally high for long decision time windows (around 85% for 60 s of data) but drops drastically for shorter decision time windows no matter which approach is used.

Recently, Wong et al. (2018) showed that decoding models outperform encoding models in terms of classification accuracies. One of the best classification results obtained so far was 85% with 20-s decision time windows, with the CCA (Geirnaert et al., 2020).

Non-linear Models

Similarly to linear models, several non-linear model architectures are in competition. But non-linear models are still overlooked because they are more complex to implement and interpret. Nevertheless, they were used by a few studies to explore AAD. Vandecapelle et al. (2020) used two convolutional neural networks to determine the attended speaker in a multi-speaker scene by using the direction of the locus of auditory attention. Their method allows them to decode auditory attention with very short decision time windows and with a good classification accuracy (around 80% for 2 s of data). In another study the authors used a fully-connected neural network to reconstruct the speech envelope and estimate the attended speaker (de Taillez et al., 2017). The classification accuracy obtained with this method appears to be similar to the performance obtained in Vandecapelle et al. (2020) even though the comparison between studies is not straightforward due to differences in experimental and model parameters or accuracy measures (Ciccarelli et al., 2019). However, non-linear models outperform linear models in terms of decision time window/performance ratio. One other potential advantage of this type of model is that it seems more realistic insofar as it may capture the neuronal non-linearity underlying speech perception (O'Sullivan et al., 2015, Mirkovic et al., 2015, de Taillez et al., 2017).

Limitations of Linear and Non-linear Models

Linear and non-linear models yet suffer from several limitations with respect to AAD. The major problem of linear models lies in the fact that their classification accuracy is strongly influenced by the duration of the decision window. Long windows yield good classification (>80%) while short ones (e.g., 2 s) yield much poorer performance (∼60%). This is due to the fact that (1) short decision windows contain less information (Vandecapelle et al., 2020), (2) EEG signals contain a mixture of several physiological and neural processes. Thus, correlations between predicted and actual data are rather weak (between 0.05 and 0.2) and short decision time windows are particularly sensitive to noise (Geirnaert et al., 2020). Moreover, a huge amount of data is needed to fit the model properly. Therefore, these models are difficult to use in real time situations where the selection of the attended speaker must be performed as fast as possible.

For non-linear models, the principal issue is the risk of overfitting, in particular with small datasets (Vandecapelle et al., 2020). Moreover, comparing performances of several non-linear models on different datasets pointed to a low reproducibility of these algorithms (Geirnaert et al., 2020). Besides fitting issues and physiological noise (and non-relevant neural signal), another source of performance variability resides in inter-individual differences at the cognitive level, such as for instance in working memory (WM) (Ciccarelli et al., 2019), attentional control, cognitive inhibition, but also motivation.

FUTURES PLAUSIBLE APPLICATIONS FOR AUDITORY ATTENTION DETECTION METHODS

Plausible Applications for AAD-Passive Brain Computer Interfaces Systems

Classical *active* Brain Computer Interfaces (*a*BCI) exploit the user's voluntary brain activity to control applications or devices. Several years ago, a new category of BCI, named *passive* Brain Computer Interfaces (*p*BCI), emerged. Unlike *a*BCI, *p*BCI use involuntary brain activity (e.g., cognitive state) to implicitly modify human-machine interactions (Zander and Kothe, 2011; Clerc et al., 2016). *passive* Brain Computer Interfaces are generally used to monitor attention, fatigue, or workload in real life situations such as driving situations (Haufe et al., 2014) or air traffic control (Aricò et al., 2016) but they can also be used in less operational contexts. For example, *p*BCI can be used to provide translation of unknown read words (Hyrskykari, 2006) or to display information on the screen when the user needs it (Jacob, 1990). *passive* Brain Computer Interfaces also have applications in the field of virtual reality and video gaming (Lécuyer et al., 2008; George and Lécuyer, 2010).

Auditory attention detection algorithms could be coupled with passive BCI to extend the usefulness of such methods to more concrete applications. In the next section, we will describe some possible future applications for AAD-*p*BCI systems.

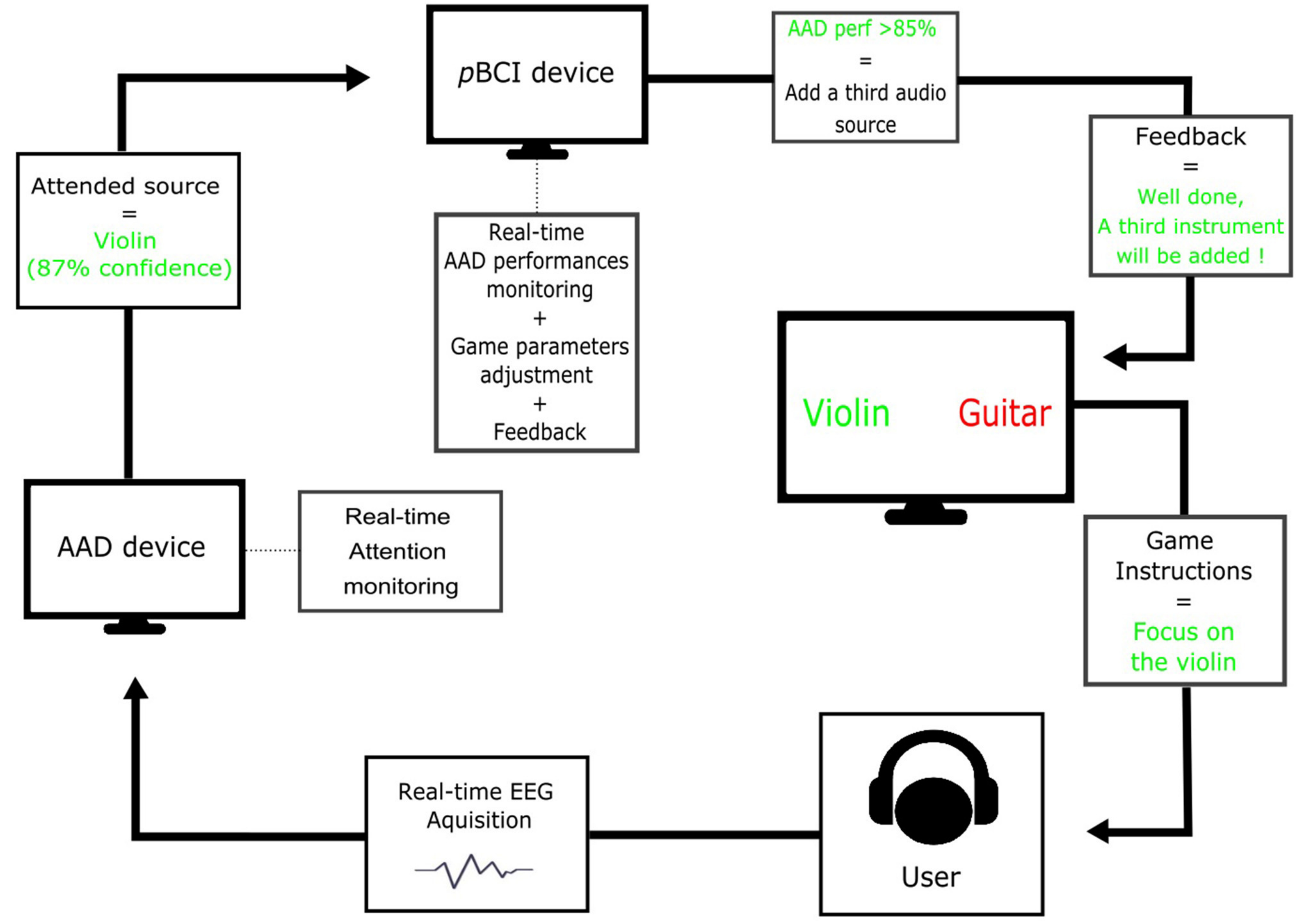


FIGURE 1 | Schematic representation of a sustained attention enhancement AAD-pBCI system based on a serious game. While the user is concentrating on a specific auditory source among several others, brain activity is recorded, and preprocessed in real time. Based on this recorded EEG data, the attended auditory source is continuously tracked by the AAD device. The pBCI device collects the AAD performances continuously (i.e., regression accuracy), estimates if a certain threshold has been exceeded, adapts, in real-time, the game parameters (e.g., instructions, auditory scene complexity), and gives feedback to the user.

AAD-pBCI in Education

Since a few years, studies that explore the relationship between children's attention abilities and screen access have shown that precocious screen access may go along with attentional problems (Christakis et al., 2004; Ponti et al., 2017; Tamana et al., 2019, but see Kostyrka-Allchorne et al., 2017 for a systematic review on the relationship between television exposure and children's cognition). AAD-*p*BCI systems could be used to improve children's attention ability. Such an attempt was made by Cho et al. (2002) who developed an attention enhancement system for ADHD children using EEG biofeedback and a virtual classroom environment. They showed that it is possible to use *p*BCI to enhance attention in children with ADHD in a school context. An advantage of real-time AAD applications is that they may allow monitoring children's attention. Moreover, they could be of use in serious game applications aiming at enhancing sustained auditory attention (see for instance **Figure 1**). Importantly, one can hypothesize that, because sustained attention in a complex auditory scene requires segregation and integration abilities but also inhibition and WM, these functions may also benefit from such applications.

Such a tool could also benefit musicians who must be able to sustain attention for long periods of time (Bergman Nutley et al., 2014). Interestingly, for musicians, this approach could also enhance the ability to share auditory attention across multiple sources, since this is of great importance in ensemble music making. As for the Sustained Attention Enhancement AAD-*p*BCI System mentioned above, a Divided Attention Enhancement AAD-*p*BCI System could also take the form of a musical serious game wherein the player has to learn to switch the focus of attention from one source to another and to share attention across multiple sources.

AAD-pBCI in Art

In the field of art, several attempts have been made to bridge EEG and BCI since the 1970s (Vidal, 1973; Rosenboom, 1977; Williams and Miranda, 2018). More recently, works have been done to develop systems to control an instrument (Arslan et al., 2006) or to generate melodies with brain signals (Wu et al., 2010;

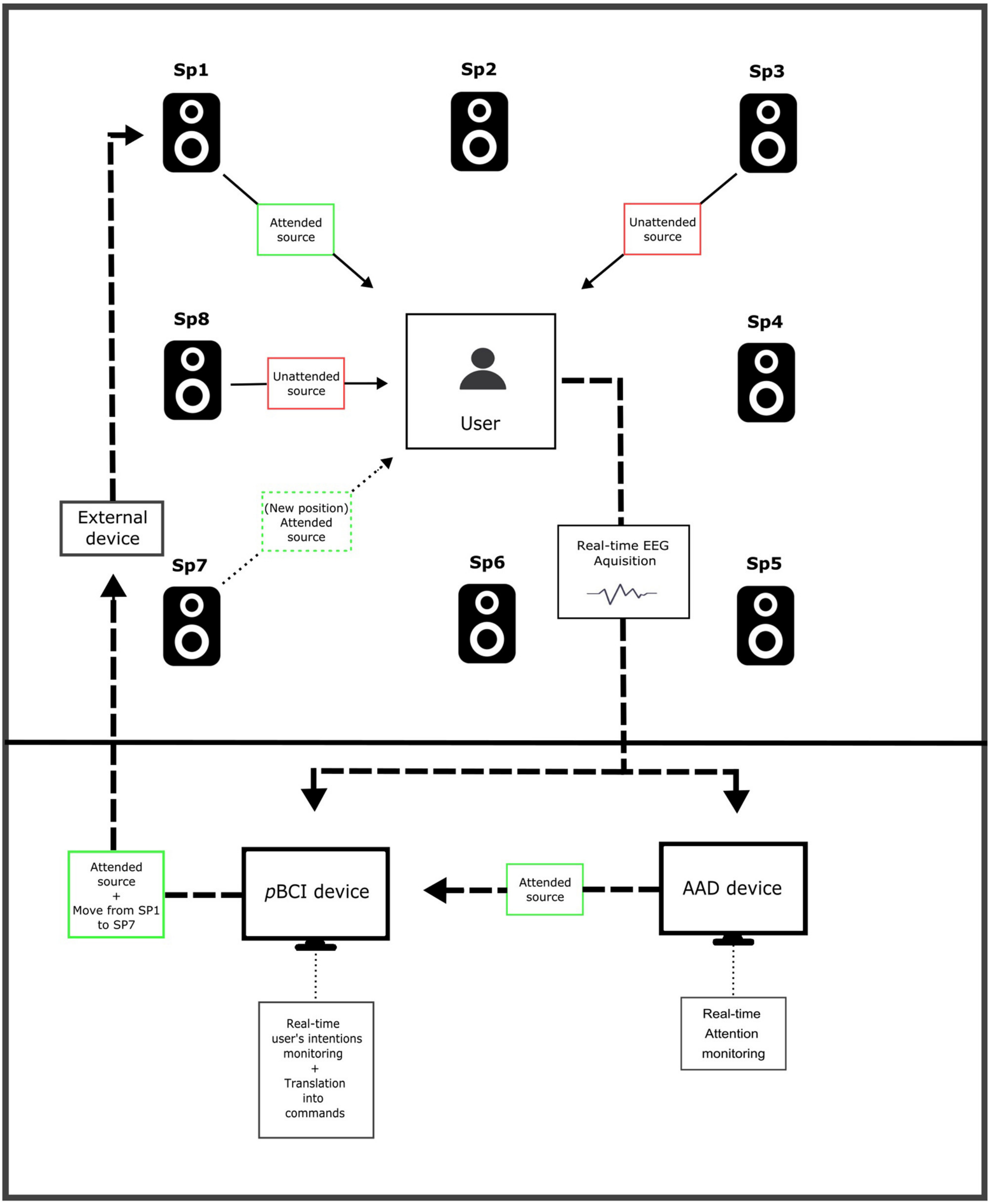


FIGURE 2 | Schematic representation of a real-time sound modulation AAD-pBCI system. Based on the real-time EEG data recording, the attended auditory source is continuously tracked by the AAD device. The pBCI device analyses in real-time the user's intentions (e.g., moving the attended source from the upper left loudspeaker to the bottom left one), translates it into commands and sends it to an external device that will modify the loudspeaker's parameters accordingly.

Miranda et al., 2011) to name a few. In this sense, there is a place for AAD-*p*BCI systems to create new kinds of art performances in which brain activity induced by auditory attention could be used to modulate different sound sources (see **Figure 2**). This could be of particular interest in an immersive listening structure composed of multiple loudspeakers (Pascal, 2020). Such a device would allow the user to select a specific sound source and modify its loudness, spatial location, or motion. In such a setup, the AAD module monitors in real-time the attended source and provides information about the source of interest to the *p*BCI module. This second module is responsible for analyzing the intentions of the user, translating them into command, and controlling an external device. To do so, the *p*BCI module classifies among several classes of neural activity induced by different cognitive processes (e.g., imaging a movement of the attended source). Once the user's intention has been detected, the *p*BCI module translates it into commands that correspond to a particular parameter's modification (e.g., moving the attended source from the upper central loudspeaker to the bottom central one) and sends them to an external device.

Application in Neuro-Steered Hearing Aids

The first reason why AAD has been investigated is to enhance hearing aids and more specifically, CI. Cochlear implant are electronic devices that allow deaf people to partly regain audition by converting audio signals to electrical signals directly stimulating the auditory nerve. While they perform well when the user is facing a unique speaker (or in quiet environment), in presence of multiple speakers performance drops dramatically because all speakers are amplified indistinctly (e.g., Zeng et al., 2008).

The solution to bypass this limitation is to inform hearing aids of the user's attentional focus. In fact, if the hearing aid was able to "know" which audio source the user is attending to, then it should be able to selectively enhance it. Therefore, combining AAD algorithms and hearing aids technologies, should lead to next-generation hearing aids allowing good performances in complex (or noisy) auditory environments (see for example: Das et al., 2016, 2020; Van Eyndhoven et al., 2017; Cantisani et al., 2020; Geirnaert et al., 2020).

Other Plausible Applications for AAD-Passive BCI Systems

One can think about other futuristic applications for AAD, in several distinct domains. For instance, in the entertainment field. It is, for example, possible to develop "auditory games" in which players, equipped with light AAD-pBCI systems, confront each other in musical battles using their auditory attention. In addition to being fun, this kind of game could be interesting to develop cognitive abilities that underlie auditory sustained attention (WM, executive control, etc.) even if it is not its main purpose. Furthermore, such a game could be adapted to a solo or a multiplayer environment.

AAD-*p*BCI systems could also find applications in the field of domotics. Indeed, a wearable AAD-*p*BCI system could be useful, in situations where ambient noise is varying constantly (e.g., in a living room), to monitor and adapt in real-time the loudness of the attended sound source (TV, hifi system, home phone, etc.).

CONCLUSION

Overall, AAD, by providing real-time cues of the auditory attentional state of an individual, opens new avenues to several applications. After a first stage of fundamental research to understand the links between auditory attention and neural signals, we are now in a second stage of applied research optimizing algorithms in terms of both classification performance and speed. In the next few years, when real-time decoding limitations will be overcome and wearable wireless systems will be developed, AAD could find applications in many domains such as education, art, health, or even domotics and online games.

AUTHOR CONTRIBUTIONS

JB and DS conceived the present article. JB wrote the manuscript and designed the figures with the support of DS and MC. All authors contributed to the article and approved the submitted version.

FUNDING

JB is funded by Oticon Medical (research agreement n°14294-CRI04) and Region Provence-Alpes-Côte d'Azur. DS is supported by APA foundation (RD-2016-9), ANR-16-CE28-0012-01 (RALP), ANR-CONV-0002 (ILCB), ANR-11-LABX-0036 (BLRI), and the Excellence Initiative of Aix-Marseille University (A*MIDEX).

REFERENCES

Akram, S., Presaccoc, A., Simon, J. Z., Shamma, S. A., and Babadi, B. (2016). Robust decoding of selective auditory attention from MEG in a competing-speaker environment via state-space modeling. *Neuroimage* 124, 906–917. doi: 10.1016/j.neuroimage.2015.09.048

Alickovic, E., Lunner, T., Gustafsson, F., and Ljung, L. (2019). A tutorial on auditory attention identification methods. *Front. Neurosci.* 13:153. doi: 10.3389/fnins.2019.00153

Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Bonelli, S., Golfetti, A., et al. (2016). Adaptive automation triggered by EEG-based mental workload index: a passive brain-computer interface application in realistic air traffic control environment. *Front. Hum. Neurosci.* 10:539. doi: 10.3389/fnhum.2016.00539

Arslan, B., Brause, A., Castet, J., Léhembre, R., Simon, C., Filatriau, J. J., et al. (2006). "A real time music synthesis environment driven with biological signals," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, Vol. 2 (Toulouse), 1172–1175. doi: 10.1109/icassp.2006.1660557

Bergman Nutley, S., 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

Bleichner, M. G., Mirkovic, B., and Debener, S. (2016). Identifying auditory attention with ear-EEG: CEEGrid versus high-density cap-EEG comparison. *J. Neural Eng.* 13:066004. doi: 10.1088/1741-2560/13/6/066004

Cantisani, G., Essid, S., and Richard, G. (2019). "EEG-based decoding of auditory attention to a target instrument in polyphonic music," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (New Paltz, NY), 80–84. doi: 10.1109/WASPAA.2019.8937219

Cantisani, G., Essid, S., Richard, G., Cantisani, G., Essid, S., Richard, G., et al. (2020). Neuro-steered music source separation with EEG-based auditory attention decoding and contrastive-NMF. *Hal Archives-Ouvertes.*

Cho, B. H., Lee, J. M., Ku, J. H., Jang, D. P., Kim, J. S., Kim, I. Y., et al. (2002). "Attention enhancement system using virtual reality and EEG biofeedback," in *Proceedings - Virtual Reality Annual International Symposium* (Seattle, WA), 156–163. doi: 10.1109/vr.2002.996518

Christakis, D. A., Zimmerman, F. J., DiGiuseppe, D. L., and McCarty, C. A. (2004). Early television exposure and subsequent attentional problems in children. *Pediatrics* 113, 708–713. doi: 10.1542/peds.113.4.708

Ciccarelli, G., Nolan, M., Perricone, J., Calamia, P. T., Haro, S., O'Sullivan, J., et al. (2019). Comparison of two-talker attention decoding from EEG with nonlinear neural networks and linear methods. *Sci. Rep.* 9, 1–10. doi: 10.1038/s41598-019-47795-0

Clerc, M., Bougrain, L., and Lotte, F. (2016). *Brain–Computer Interfaces 1: Foundations and Methods*. New York, NY: Wiley.

Crosse, M. J., Di Liberto, G. M., Bednar, A., and Lalor, E. C. (2016). The multivariate temporal response function (mTRF) toolbox: a MATLAB toolbox for relating neural signals to continuous stimuli. *Front. Hum. Neurosci.* 10:604. doi: 10.3389/fnhum.2016.00604

Das, N., Bertrand, A., and Francart, T. (2018). EEG-based auditory attention detection: boundary conditions for background noise and speaker positions. *BioRxiv* 32, 1–18. doi: 10.1101/312827

Das, N., Van Eyndhoven, S., Francart, T., and Bertrand, A. (2016). "Adaptive attention-driven speech enhancement for EEG-informed hearing prostheses," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)* (IEEE), 77–80. doi: 10.1109/EMBC.2016.7590644

Das, N., Zegers, J., Van Hamme, H., Francart, T., and Bertrand, A. (2020). Linear versus deep learning methods for noisy speech separation for EEG-informed attention decoding. *J. Neural Eng.* 17:046039. doi: 10.1088/1741-2552/aba6f8

de Cheveigné, A., Wong, D. E., Di Liberto, G. M., Hjortkjær, J., Slaney, M., and Lalor, E. (2018). Decoding the auditory brain with canonical component analysis. *NeuroImage* 172, 206–216. doi: 10.1016/j.neuroimage.2018.01.033

de Taillez, T., Kollmeier, B., and Meyer, B. T. (2017). Machine learning for decoding listeners' attention from EEG evoked by continuous speech. *Eur. J. Neurosci.* 51, 1234–1241. doi: 10.1111/ijlh.12426

Ding, N., and Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11854–11859. doi: 10.1073/pnas.1205381109

Fiedler, L., Wöstmann, M., Graversen, C., Brandmeyer, A., Lunner, T., and Obleser, J. (2017). Single-channel in-ear-EEG detects the focus of auditory attention to concurrent tone streams and mixed speech. *J. Neural Eng.* 14:036020. doi: 10.1088/1741-2552/aa66dd

Fuglsang, S. A., Dau, T., and Hjortkjær, J. (2017). Noise-robust cortical tracking of attended speech in real-world acoustic scenes. *NeuroImage* 156, 435–444. doi: 10.1016/j.neuroimage.2017.04.026

Geirnaert, S., Vandecappelle, S., Alickovic, E., de Cheveigné, A., Lalor, E., Meyer, B. T., et al. (2020). Neuro-steered hearing devices: decoding auditory attention from the brain. *ArXiv* 802895, 1–20. Available online at: http://arxiv.org/abs/2008.04569

George, L., and Lécuyer, A. (2010). "An overview of research on "passive" brain-computer interfaces for implicit human-computer interaction," in *International Conference on Applied Bionics and Biomechanics ICABB 2010 - Workshop W1 "Brain-Computer Interfacing and Virtual Reality"* (Venise). Available online at: http://hal.inria.fr/docs/00/53/72/11/PDF/GeorgeL-LecuyerA.pdf

Giraud, A. L., and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15, 511–517. doi: 10.1038/nn.3063

Haufe, S., Kim, J. W., Kim, I. H., Sonnleitner, A., Schrauf, M., Curio, G., et al. (2014). Electrophysiology-based detection of emergency braking intention in real-world driving. *J. Neural Eng.* 11:056011. doi: 10.1088/1741-2560/11/5/056011

Holdgraf, C. R., Rieger, J. W., Micheli, C., Martin, S., Knight, R. T., and Theunissen, F. E. (2017). Encoding and decoding models in cognitive electrophysiology. *Front. Syst. Neurosci.* 11:61. doi: 10.3389/fnsys.2017.00061

Hyrskykari, A. (2006). *Eyes in Attentive Interfaces: Experiences from Creating iDict, A Gaze-Aware Reading Aid*. (Issue January 2006). Available online at: http://acta.uta.fi

Jacob, R. J. K. (1990). "What you look at is what you get: eye movement-based interaction techniques," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Montréal, QC), 11–18.

Kostyrka-Allchorne, K., Cooper, N. R., and Simpson, A. (2017). The relationship between television exposure and children's cognition and behaviour: a systematic review. *Dev. Rev.* 44, 19–58. doi: 10.1016/j.dr.2016.12.002

Kubanek, J., Brunner, P., Gunduz, A., Poeppel, D., and Schalk, G. (2013). The tracking of speech envelope in the human cortex. *PLoS ONE* 8:e53398. doi: 10.1371/journal.pone.0053398

Lécuyer, A., Lotte, F., Reilly, R. B., and College, T. (2008). Brain-computer interfaces, virtual reality, and videogames. *Computer* 41, 66–72. doi: 10.1109/MC.2008.410

Mesgarani, N., and Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* 485, 233–236. doi: 10.1038/nature11020

Mesgarani, N., David, S. V., Fritz, J. B., and Shamma, S. A. (2009). Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. *J. Neurophysiol.* 102, 3329–3339. doi: 10.1152/jn.91128.2008

Miran, S., Akram, S., Sheikhattar, A., Simon, J. Z., Zhang, T., and Babadi, B. (2018a). Real-time tracking of selective auditory attention from M/EEG: a Bayesian filtering approach. *Front. Neurosci.* 12:262. doi: 10.3389/fnins.2018.00262

Miran, S., Akram, S., Sheikhattar, A., Simon, J. Z., Zhang, T., and Babadi, B. (2018b). "Real-time decoding of auditory attention from EEG via Bayesian filtering," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* (New Orleans, LA), 25–28. doi: 10.1109/EMBC.2018.8512210

Miranda, E. R., Magee, W. L., Wilson, J. J., Eaton, J., and Palaniappan, R. (2011). Brain-Computer Music Interfacing (BCMI): from basic research to the real world of special needs. *Music Med.* 3, 134–140. doi: 10.1177/1943862111399290

Mirkovic, B., Debener, S., Jaeger, M., and De Vos, M. (2015). Decoding the attended speech stream with multi-channel EEG: implications for online, daily-life applications. *J. Neural Eng.* 12:046007. doi: 10.1088/1741-2560/12/4/046007

Nourski, K. V., Reale, R. A., Oya, H., Kawasaki, H., Kovach, C. K., Chen, H., et al. (2009). Temporal envelope of time-compressed speech represented in the human auditory cortex. *J. Neurosci.* 29, 15564–15574. doi: 10.1523/JNEUROSCI.3065-09.2009

O'Sullivan, J., Chen, Z., Sheth, S. A., McKhann, G., Mehta, A. D., and Mesgarani, N. (2017). "Neural decoding of attentional selection in multi-speaker environments without access to separated sources," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* (Jeju), 1644–1647. doi: 10.1109/EMBC.2017.8037155

O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., et al. (2015). Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cereb. Cortex* 25, 1697–1706. doi: 10.1093/cercor/bht355

Pascal, M. (2020). Analyse de la composition de l'espace dans une œuvre acousmatique immersive de Jean Marc Duchenne. *Hal Archives-Ouvertes.* 2020:hal-02926984.

Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., et al. (2012). Reconstructing speech from human auditory cortex. *PLoS Biol.* 10:e1001251. doi: 10.1371/journal.pbio.1001251

Ponti, M., Bélanger, S., Grimes, R., Heard, J., Johnson, M., Moreau, E., et al. (2017). Screen time and young children: Promoting health and development in a digital world. *Paediatr. Child Health.* 22, 461–477. doi: 10.1093/pch/pxx123

Rosenboom, D. (1977). Biofeedback and the arts: results of early experiments. *J. Aesthet. Art Critic.* 35, 385–386.

Schäfer, P. J., Corona-Strauss, F. I., Hannemann, R., Hillyard, S. A., and Strauss, D. J. (2018). Testing the limits of the stimulus reconstruction approach: auditory attention decoding in a four-speaker free field environment. *Trends Hear.* 22, 1–12. doi: 10.1177/2331216518816600

Tamana, S. K., Victor, E., Joyce, C., Diana, L. L., Meghan, B., A., et al. (2019). Screen-time is associated with inattention problems in preschoolers: results from the CHILD birth cohort study. *PLoS ONE* 14:e0213995. doi: 10.1371/journal.pone.0213995

Van Eyndhoven, S., Francart, T., and Bertrand, A. (2017). EEG-informed attended speaker extraction from recorded speech mixtures with application in neuro-steered hearing prostheses. *IEEE Trans. Biomed. Eng.* 64, 1045–1056. doi: 10.1109/TBME.2016.2587382

Vandecapelle, S., Deckers, L., Das, N., Ansari, A. H., Bertrand, A., and Francart, T. (2020). EEG-based detection of the attended speaker and the locus of auditory attention with convolutional neural networks. *BioRxiv [Preprint]*. doi: 10.1101/475673

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

Williams, D., and Miranda, E. R. (2018). "BCI for music making: then, now, and next," in *Brain–Computer Interfaces Handbook: Technological and Theoretical Advances*, eds C. S. Nam, A. Nijholt, and F.Lotte (Florence, KY: CRC Press), 191–205.

Wong, D. D. E., Fuglsang, S. A., Hjortkjær, J., Ceolini, E., Slaney, M., and de Cheveigné, A. (2018). A comparison of regularization methods in forward and backward models for auditory attention decoding. *Front. Neurosci.* 12:531. doi: 10.3389/fnins.2018.00531

Wu, D., Li, C., Yin, Y., Zhou, C., and Yao, D. (2010). Music composition from the brain signal: Representing the mental state by music. *Comput. Intell. Neurosci.* 2010:267671. doi: 10.1155/2010/267671

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.* 8:025005. doi: 10.1088/1741-2560/8/2/025005

Zeng, F. G., Rebscher, S., Harrison, W., Sun, X., and Feng, H. (2008). Cochlear implants: system design, integration, and evaluation. *IEEE Rev. Biomed. Eng.* 1, 115–142. doi: 10.1109/RBME.2008.2008250

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., et al. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party." *Neuron* 77, 980–991. doi: 10.1016/j.neuron.2012.12.037

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

frontiers in Computer Science

ORIGINAL RESEARCH
published: 07 May 2021
doi: 10.3389/fcomp.2021.661466

Exploring the Brain Activity Related to Missing Penalty Kicks: An fNIRS Study

Max W. J. Slutter, Nattapong Thammasan and Mannes Poel*

Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, Netherlands

At vital moments in professional soccer matches, penalties were often missed. Psychological factors, such as anxiety and pressure, are among the critical causes of the mistakes, commonly known as *choking under pressure*. Nevertheless, the factors have not been fully explored. In this study, we used functional near-infrared spectroscopy (fNIRS) to investigate the influence of the brain on this process. An *in-situ* study was set-up ($N = 22$), in which each participant took 15 penalties under three different pressure conditions: without a goalkeeper, with an amiable goalkeeper, and with a competitive goalkeeper. Both experienced and inexperienced soccer players were recruited, and the brain activation was compared across groups. Besides, fNIRS activation was compared between sessions that participants felt anxious against sessions without anxiety report, and between penalty-scoring and -missing sessions. The results show that the task-relevant brain region, the motor cortex, was more activated when players were not experiencing performance anxiety. The activation of task-irrelevant areas was shown to be related to players experiencing anxiety and missing penalties, especially the prefrontal cortex (PFC). More particularly, an overall higher activation of the PFC and an increase of PFC lateral asymmetry were related to anxious players and missed penalties, which can be caused by players' worries about the consequences of scoring or missing the penalty kicks. When experienced players were feeling anxious, their left temporal cortex activation increased, which could be an indication that experienced overthink the situation and neglect their automated skills. Besides, the left temporal cortex activation is higher when inexperienced players succeeded to score a penalty. Overall, the results of this study are in line with the neural efficiency theory and demonstrate the feasibility and ecological validity to detect neurological clues relevant to anxiety and performance from fNIRS recordings *in the field*.

Keywords: fNIRS, football, soccer, sports, penalty kick, choking, mental pressure, neural efficiency

OPEN ACCESS

Edited by:
Aleksander Väljamäe,
Tallinn University, Estonia

Reviewed by:
Muhammad Jawad Khan,
National University of Sciences and Technology (NUST), Pakistan
Antonio R. Hidalgo-Muñoz,
Université de Toulouse, France

****Correspondence:***
Nattapong Thammasan
n.thammasan@utwente.nl

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Computer Science

Received: *30 January 2021*
Accepted: *06 April 2021*
Published: *07 May 2021*

Citation:
Slutter MWJ, Thammasan N and Poel M (2021) Exploring the Brain Activity Related to Missing Penalty Kicks: An fNIRS Study. Front. Comput. Sci. 3:661466. doi: 10.3389/fcomp.2021.661466

1. INTRODUCTION

Penalty kicks are highly important in soccer. Penalties are common and have a big influence on the outcome of a match. By taking the large amounts of money and number of fans into account, the importance of penalty kicks increases even more. In other words, missing a penalty in a crucial match can cause thousands of fans to be disappointed and the corresponding club to miss out on millions of euros. Many technical skills have an influence on the quality of the penalty kick (see

Memmert et al., 2013 for review). Apart from technical skill, psychological factors seem to have a clear influence on the outcome of a penalty kick as well. It was found that only psychological factors had a large negative influence on the outcome of the penalty, where skill and fatigue did not (Jordet et al., 2007).

Many studies on the causes of missed penalties have convincingly shown that the kicker's anxiety and the mental pressure under which the kicker are the most common psychological factors. While resistance to mental pressure depends on player personality (Lin et al., 2017), the pressure often leads to distress, which is a negative factor adversely influencing the quality of the penalty kick and thereby hindering scoring, rather than eustress that is positive and gives a feeling of arousal and thereby enhancing chance of scoring (Le Fevre et al., 2003). The degraded performance under pressure and anxiety is often referred to as *choking*, which is prevailing in the critical moments of the soccer big matches (Chiappori et al., 2002; Arrondel et al., 2019). Anxiety, as a result of choking under pressure, was found to be related to bad direction of penalty taking (Wilson et al., 2009), and this adverse effect is also present in other sports domain, such as weightlifting (Genakos and Pagliero, 2011), golf (Hickman and Metz, 2015), chess (González-Díaz and Palacios-Huerta, 2016), basketball (Fryer et al., 2018), and tennis (Cohen-Zada et al., 2017).

Choking under pressure is generally explained by self-focusing theory or distraction theory. The self-focus theory posits that anxiety or pressure increases the level of self-consciousness, resulting in more consciously monitoring or controlling skill execution, and choking as a result (Baumeister, 1984; Hill et al., 2010; Roberts et al., 2017). This means that excessive pressure leads to the undermining of automatism and therefore there is overwhelmed attention toward the execution of the skill. On the other hand, the distraction-theory posits that anxiety or pressure occupies the working memory, causing a shift from task-relevant cues to task-irrelevant cues (Sarason, 1988; Hill et al., 2010; Gröpel and Mesagno, 2017; Roberts et al., 2017). Unlike the self-focus theory, little attention is paid toward the execution of the skill, where the distractions can be either internal (e.g.,, worries) or external (e.g., distracting fans), which can be explained by the circles of attention (Eberspächer et al., 1990). The two theory are relevant to the *neural efficiency* theory, positing that expert athletes show more efficient brain activity than non-athletes, meaning that task-relevant activities are increased and task-irrelevant activities are decreased. A task-irrelevant activity, such as planning and worries about thinking about consequences of missing penalties, can be a distracting factor suppressing task-relevant activities, such as motor controlling. These two theories of choking can be connected via the fear circuit model (Hatfield and Kerick, 2007), which involves the prefrontal cortex (PFC), basal ganglia, thalamus, premotor cortex, motor cortex, limbic system, anterior cingulate cortex, left temporal cortex, and the corticospinal tract.

Many human neuroimaging studies have provided neural evidence of choking in the brain above and beyond behavioral analysis. First, increased PFC activation will act as a distraction in the brain leading to choking. The study of Korb (2010) suggested that an increase in PFC activation is associated with being distracted from a physical task and with being stressed, although the opposite trend was observed in the study of Al-shargie et al. (2016) in a different task. Second, a decrease in PFC lateral asymmetry will lead to choking (Hatfield and Kerick, 2007), where the improved performance was found associated with higher left compared to right PFC activation (Silveira et al., 2019). Third, the neural efficiency theory suggests that optimal performance can be achieved by activating task-relevant brain areas. This means an increase in motor cortex activity is associated with being less likely to choke when performing a sport-related exercise, and supportive evidence was reported in a study of Wolf et al. (2014) in expert table tennis players. Forth, intensive connectivity between dorsolateral PFC (DLPFC) and the motor cortex was found necessary for maintaining the level of performance in motor-related tasks, thus preventing choking (Yoon et al., 2006; Clapp et al., 2009; Lee and Grafton, 2015). Last, a heightened left temporal cortex activation was found associated with choking in experts (Wolf et al., 2015) due to self-instruction, suppressed automated skill, disturbing thoughts, similar to the phenomenon found in novice players. This is supported by the reported lower activation in this brain region of experts in shooting sports (Haufler et al., 2000, 2002; Allen et al., 2004; Kerick et al., 2004) due to lower cognitive demands.

Despite the neurological evidence of choking under pressure, the study of this phenomenon *in the field* is limited. In particular, there can be a huge difference between controlled-laboratory-setting choking, in which simple sensorimotor tasks [such as feet tapping, treadmill walking, and smartphone-based touch game (Udina et al., 2020)] are performed, and the real-life naturalistic-setting choking, where a wide range of external factors involve. A possible reason for the lack of choking study *in the field* is the susceptibility to movement artifacts of neuroimaging techniques. Recently, functional near-infrared spectroscopy (fNIRS) has been often used in *in situ* studies of brain activity due to its compelling robustness to movement artifacts, which has been proven in the study of Carius et al. (2020). In that study, brain activity was reliably measured during bouldering—a special form of climbing without a rope where complex whole-body movements are involved, and it was demonstrated that fNIRS is capable of measuring sensorimotor activity during the execution of heavy and irregular movements. As bouldering involves more strenuous movements than kicking a ball, it should be possible to measure brain activity in the soccer domain as well, which has not been fully explored yet. To our best knowledge, there has been only one fNIRS study in penalty kick of soccer, which compared the brain activity of experienced and inexperienced goalkeepers who were instructed to watch pre-recorded videos of penalty kicks from the perspective of goalkeepers (Kuriyama et al., 2015). However, an actual physical movement was not involved and choking was not focused. On the other hand, effects of pressure on poorer performance have been demonstrated in a previous work (Ito et al., 2011), but its working-memory task did not include physical activity. To date, the feasibility to capture choking effect *in the field* using fNIRS is still unclear.

In this study, we aimed to explore brain activity during the real situation of penalty kicking using fNIRS. Crucially, we sought to examine the left temporal cortex, motor cortex, PFC, and functional connectivity between DLPFC and motor cortex during choking to characterize brain activity that involves anxiety and impacts performance. The findings of this study can have an implication on a wider range of tasks beyond the soccer/sports domain, such as in surgery, where motor performance under high mental pressure involves. This study provides insights on why people fail to perform under pressure and possibly paves a way toward tailored intervention to prevent choking by utilizing a closed-loop brain-computer interface.

The present study also aimed to investigate the correlation between the level of expertise and the capability to deal with pressure. It was evidenced that the brain activity of sports professionals differ by level of expertise (Kuriyama et al., 2015; Wolf et al., 2015), where experts show more efficient brain activity or activate the *correct* areas of the brain for a certain activity when performing a skill. Under mental pressure, we predicted that experience in sports might also influence the way to cope with anxiety, leading to different patterns in brain activity between experts and novices during choking.

In general, previous works leave open the critical question of to what extent brain activity associated with choking under pressure in a penalty kick situation can be reflected by *in-the-field* fNIRS measurement. While theories of choking under pressure are under development, we focused more on the anxiety and pressure (Yu, 2015), which are the associated psychological factors that can be explicitly measured and strongly induced using established methods in sports psychology. Specifically, we formulated our research questions as follows:

RQ1: Performance—What is the difference in brain activity between performing success (scoring) and failure (missing) when taking a penalty kick?

RQ2: Performing under pressure—What brain activity is associated with performing under pressure during a penalty kick situation?

RQ3: Experienced and Inexperienced players—What is the general difference in brain activity between experienced and inexperienced soccer players when taking a penalty kick?

RQ4: Anxiety and Experienced players—What brain activity is associated with experienced soccer players that experience (performance) anxiety when taking a penalty kick?

RQ5: Anxiety and Inexperienced players—What brain activity is associated with inexperienced soccer players that experience (performance) anxiety when taking a penalty kick?

2. MATERIALS AND METHODS

2.1. Participants

In total, 22 participants (10 females; age average: 22.9 years, standard deviation: 2.00 years) were recruited to participate in the experiment. Among these, ten participants were experienced soccer players who were in the first team of vv Drienerlo—the soccer association of the University of Twente—and trained and played matches regularly. On the other hand, 12 inexperienced participants who were also recruited never played or had limited experience in soccer. The short form of Edinburgh Handedness Inventory (Veale, 2013) was used to confirm that all participants were right-footed and right-handed, with an average Laterality Quotient of 77.27. The Sport Competition Anxiety Test (SCAT), consisting of 15 items in the range of 10–30, was used to indicate the level of performance anxiety of the participants (Martens, 1977; Martens et al., 1990; Wood, 2017a) before the experiment; the results indicate that eight, 11, and three of participants were classified into the group with low (score <17), medium (score 17–24), and high (score >24) performance anxiety, respectively, with the grand average score of 18.3. All participants provided written informed consent to participate in the study.

2.2. Tasks and Procedure

The experiment has been approved by the ethics committee of the EEMCS faculty of the University of Twente (reference number: RP 2020-118). During the experiment, the participants were instructed to perform a penalty kicking task for three rounds, each of which consisted of five penalties (trials). While pressure induction level differs by round, the same rules applied for each penalty (trial); before the penalty can be taken, the player had to wait for the referee to blow the whistle. The goalkeeper had to stay on the goal line until the ball was struck. However, the goalkeeper was allowed to move horizontally on the goal line. The player was not allowed to pause (fully stand still) during the run-up but was allowed to slow down in order to trick the goalkeeper.

The experiment consisted of three rounds and the aim was to increase the pressure per round. In the first round, the lowest amount of pressure should be induced and in the last round, the highest pressure should be induced. The specifics of each round were set-up in cooperation with a sports psychologist of the NOC*NSF (the Dutch overarching sports organization), who interacts directly with sports professionals. Based on the advice and expertise of the sports psychologist, three rounds were set-up as follows:

1. No goalkeeper: during the first round, no goalkeeper was present. The player was shooting at an empty goal and was informed that as it was a practice round to familiarize the player with the experimental protocol. Therefore, low pressure was expected from this round.
2. Amiable goalkeeper: during the second round, a goalkeeper was present but was not allowed to distract the player, who was informed that it was a friendly competition between the player and the goalkeeper and that this round aimed to see how well the player could perform against a goalkeeper. Neither the goalkeeper nor the researcher was allowed to respond to the performance of the player. By introducing a non-interacting goalkeeper, the aim was to introduce the competitive element, without raising the pressure too much.
3. Competitive goalkeeper: during the last round, we aimed to maximize the mental pressure of the player. A goalkeeper was present and allowed to distract the player, who was informed that it was a competition in which only the best performing experienced and inexperienced players could win two 50-euro giftcards, assessed by the number of goals

scored and the quality of penalty taking. To imitate a real-life professional penalty shoot-out, the player was instructed to start from the halfway line (about 40–50 m from the goal), walk with the ball toward the penalty spot, and place the ball on the right spot. This would prolong the time of being anxious, in which the researcher also tried to involve in pressure induction by providing unrealistically good statistics of previous participants or the confronting goalkeeper. Besides, the goalkeeper also tried to intimidate the player by awaiting the player at the penalty spot to which the player was approaching from the halfway line, wasting time (by drinking water or retying their shoelaces), talking to the player when he/she tried to concentrate, repeatedly calling the player by the first name, stretching arms, jumping, and telling the player that he/she already knew the direction of the upcoming shooting. The aim of these actions of the goalkeeper and researcher was to shift the attention of the players from their task (Eberspächer et al., 1990).

When the participants had finished the SCAT questionnaire, the fNIRS headset was attached to the participants. Whilst the researcher was verifying the quality of each channel, the structure of the experiment was explained to the participants. An overview of the experimental protocol is depicted in **Figure 2**. For every round, a resting period of 30 s was recorded first. During this period, the participants were instructed to refrain from moving and speaking but to keep their eyes open and to look in one certain direction, preferably where they could see as little distracting external stimuli as possible.

An explanation of the round followed. It was chosen to explain the details of each round after the resting period to ensure that the participants were not thinking/worrying about the upcoming round during the resting period. After the round was explained, the participants were briefly interviewed by asking how confident they were and how many goals they thought they would score. The interview is expected to help assess how anxious the players were. After placing the ball on the penalty spot and preparing for the run-up, the participants were instructed to wait for 5 s, until the researcher indicated that they could kick the ball. The researcher was tracking the time by using the built-in stopwatch of the OxySoft software (used for the fNIRS measurement). Using this software, markers were placed during the experiment to indicate the start and end of each 5-s waiting period. These 5-s periods were used for the data analysis, as the player was standing still, minimizing the chances of motion artifacts. The participants were also instructed to minimize body movement during this period. This 5-s waiting period was included before every kick. When all five penalties were taken, the participants were asked to fill out a small questionnaire. This questionnaire included two questions in a five-point Likert scale, regarding the satisfaction with the performance and the level of motivation during that round. Furthermore, the Sport Anxiety Scale (SAS) (Smith et al., 1990, 2006; Wood, 2017b), consisting of 21 questions in a four-point Likert scale, was included to determine the level of anxiety/pressure during the round. The results of this questionnaire were used to determine whether a player was anxious or not. The next round started when the participants finished filling the questionnaire of the previous round. The fNIRS headset was equipped until all three rounds were completed. After the experiment was finished, the participants were debriefed and a structural post-interview on the experience concluded the experiment.

Whilst conducting the experiments, the majority of conditions were kept constant. An artificial soccer pitch was used to ensure the quality of the pitch was constant across experiments. Furthermore, all participants were right-handed/footed and of similar age. Every participant faced a goalkeeper of the same gender and all goalkeepers were of similar skill level, as they all played in the first team of vv Drienerlo. For all experiments, the same ball was used, namely, a Derbystar size 5, which is typically used in professional soccer matches. The air pressure of the ball was between 0.7 and 0.9 bar, following the professional soccer guidelines. As the experiment was conducted outdoor, there were also a few conditions that were variable, such as weather, temperatures which varied between 12 and 31°C, and wind force which varied from level 0 (calm) to 4 (moderate breeze) on the Beaufort scale. During three experiments, there was fog. The experiments were conducted before regular training sessions of the football club and scheduled between 4 and 8 p.m. (Central European summer time), meaning that some experiments were conducted after sunset. During these experiments, the light poles of the soccer pitch were lit. The lights were either off or on throughout the experiment, and we ensured that there was no case that the lights were switched on/off during the experiment. Therefore, ambient light was assumed consistent.

2.3. Data Acquisition

For fNIRS measurements, the Artinis Brite 24 was used to record oxygenated hemoglobin (O2Hb) and deoxygenated hemoglobin (HHb) in each channel at a sampling rate of 10 Hz. O2Hb is the form of hemoglobin with the oxygen bound, whereas HHb does not have this bound to oxygen. The Brite 24 is a portable and wireless device that allows flexibility in fNIRS optode placement with a total of 10 transmitter optodes and 8 receiver optodes. Numerous templates are available to arrange these optodes. OxySoft, which is proprietary software developed by Artinis, was used to record and transform fNIRS signals, which were then analyzed in Python. A maximum distance of 30 mm was used between each pair of optodes and a differential pathlength factor of 6 was used for all participants. During the experiment, fNIRS data were obtained from the left PFC, right PFC, left temporal cortex, motor cortex, left DLPFC, and right DLPFC, as these regions were found relevant to choking under pressure in the literature (see section 1). In order to measure all of these areas, The standard "4 × 4 + 2" template of Artinis was used[1] where the corresponding optode placement can be found in **Figure 1**. Four channels were used to record each region of the left PFC, right PFC, left temporal cortex,

[1]Artinis Medical Systems Product Catalog 2018: https://www.artinis.com/downloads.

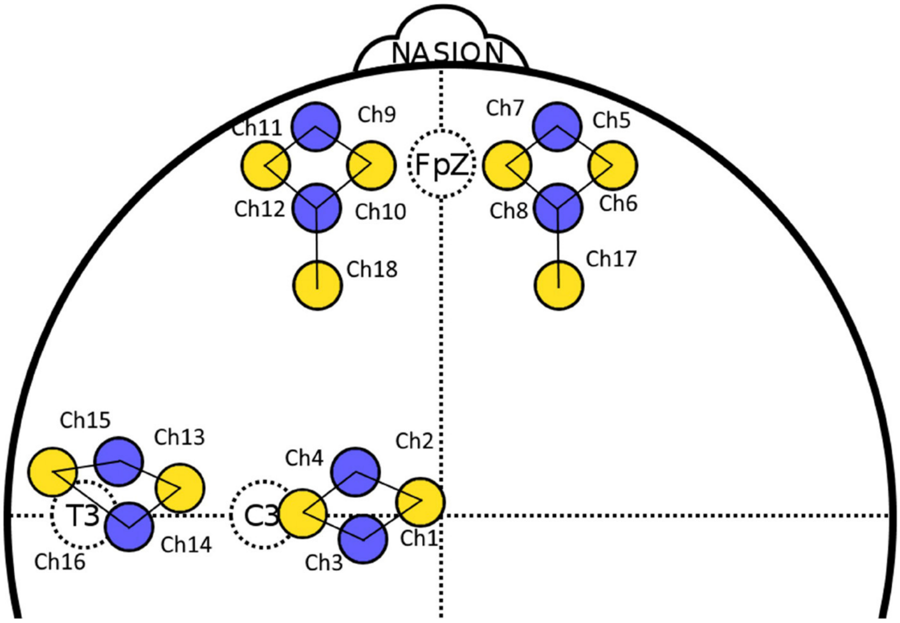


FIGURE 1 | The layout of all fNIRS channels on the scalp. The yellow circles represent transmitter optodes and the blue circles represent receiver optodes. A channel is lying between each transmitter-receiver pair. Channels 1–4 correspond to the motor cortex, channels 5–8 correspond to the right PFC, channels 9–12 correspond to the left PFC, and channels 13–16 correspond to the left temporal cortex. Channels 17 and 18 correspond to the right and left DLPFC, respectively. Certain electroencephalogram's electrode positions, in accordance with 10–20 international system, are included for references.

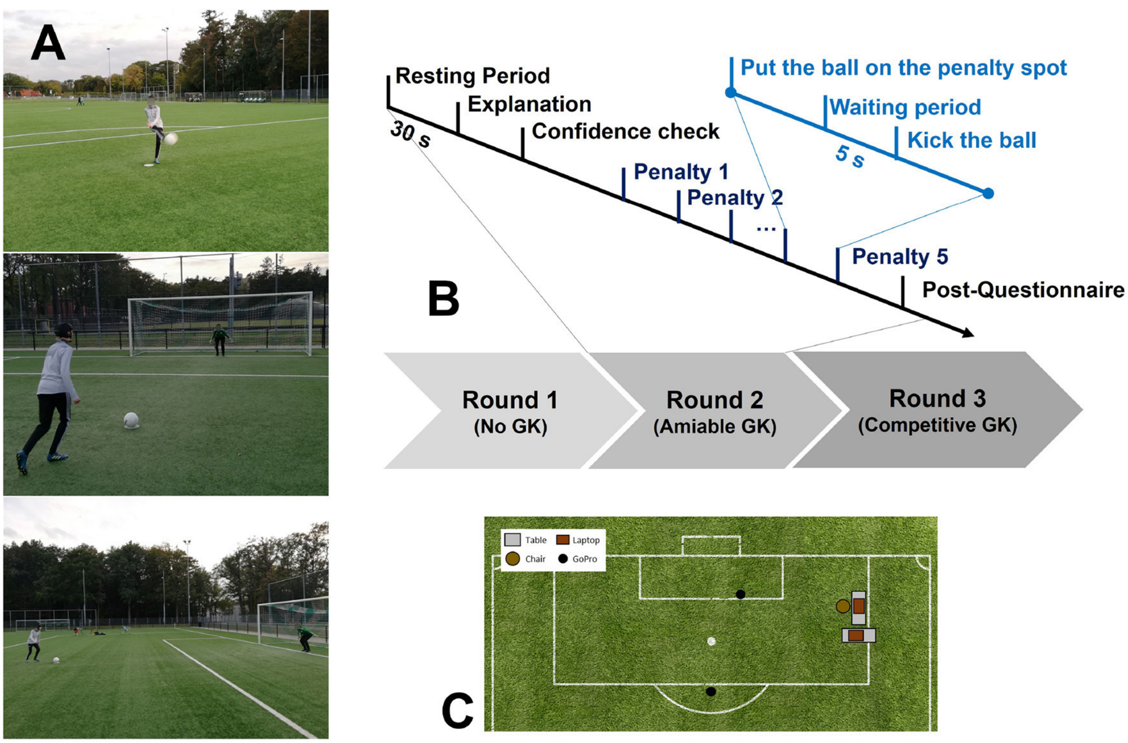


FIGURE 2 | The experimental setting: **(A)** pictures of the experiment from the front (top), back (middle), and side (bottom) angle; **(B)** experimental protocol; **(C)** placement of the equipment. The laptop close to the chair was used by the participants to fill in the questionnaires. The GoPro camera that was closest to the goal was aiming at the player and the other camera was aiming at the goal.

and motor cortex. Two channels were used to cover left and right DLPFC.

One HERO7 GoPro[2] camera was used to record the penalty kicking, such that the placement and power of the shot could be determined. The power of the shot was defined by the time it took for the ball to reach the goal. This was manually timed using a stopwatch and was expected to provide insights on kicking performance. Another GoPro camera was used to record videos of the player. We specially investigated the duration for which players looked at the goalkeeper by comparing between rounds. A longer fixation at the goalkeeper can indicate that the goalkeeper is a distracting factor (Wilson et al., 2009; Wood and Wilson, 2010; Furley et al., 2017). For consistency reasons, only the fixations during the 5-s waiting period were used. The videos of both cameras were recorded at 60 frames/s and had an image quality of 1080p.

In **Figure 2**, an overview of the set-up of the experiment is shown. The participant filled in the questionnaires on a laptop and the researcher was monitoring the fNIRS signals on a separate laptop. A Sena UD100 Bluetooth adapter[3] was used, which allowed the measurements up to a distance of 300 m. This means that the laptops could be placed at a safe distance from the goal. Furthermore, two GoPro's were used to record the player and the goal.

2.4. Data Analysis

2.4.1. Signal Pre-processing

The acquired fNIRS signals were first preprocessed by applying a fifth-order Butterworth bandpass filter between 0.02 and 0.5 Hz to get rid of physiological noises and drift in optical data (Kamran et al., 2018). The motion correction method Temporal Derivative Distribution Repair (Fishburn et al., 2019) was used to reduce the impact of motion artifacts on the signals. This novel artifact correction method shows superior performance compared to other correction methods, such as Targeted Principle Component Analysis (tPCA) (Yücel et al., 2014), correlation-based signal improvement (CBSI) (Cui et al., 2010), Movement Artifact Reduction Algorithm (MARA) (Scholkmann et al., 2010), and wavelet based methods (Molavi and Dumont, 2012; Chiarelli et al., 2015) (see Jahani et al., 2018 for a review on traditional artifact correction methods). Furthermore, TDDR method requires no parameter tuning and only minimal assumptions need to be made on the fNIRS data, while other methods assume normal distribution on fNIRS data (Cui et al., 2010) or require extensive parameter supplies from users (Scholkmann et al., 2010; Yücel et al., 2014) or suffer from baseline shift of signals (Molavi and Dumont, 2012; Chiarelli et al., 2015). The TDDR method was applied for each channel separately, using the following protocol. Given that x_t represents a datapoint of the fNIRS channel for a certain timepoint (t), the temporal derivative, y_t, of the channel was first computed by subtracting the data of the previous timepoint from the current datapoint: $y_t = x_t - x_{t-1}$.

Then, A vector of observation weight (w) was initialized: $w_t = 1$, and the weighted mean of the fluctuations (μ) was estimated by:

$$\mu = \frac{1}{\sum(w)} \sum(w_t y_t) \tag{1}$$

Afterwards the absolute residuals (r_t) of the estimated mean were computed using: $r_t = |y_t - \mu|$. An estimate of the standard deviation (σ) of these residuals was computed. This was done by multiplying the median absolute residual by the appropriate constant for the normal distribution: $\sigma = 1.4826 * median(r)$. For each observation the scaled deviation (d_t) was computed. This was done by using the standard deviation of the residuals and the tuning constant that achieves 95% efficiency on normally distributed data:

$$d_t = \frac{r_t}{4.685\sigma} \tag{2}$$

Tukey's biweight function was used to computed new observation weights:

$$w_t = \begin{cases} (1 - d_t^2)^2 & \text{if } d_t < 1 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

The steps from Equations (1) to (3) were repeated until μ converged. This was considered the case when the differences between the current μ and the previous μ was smaller than 10^{-50}. If this criterion was not satisfied after 1,000 loops (where one loop is one repetition of Equations 1–3), the process was stopped. On average 98.75 loops were needed in this process. After μ was converged, the resulting robust weights were applied to the centered temporal derivative (subtracting the mean), in order to produce the corrected derivative ($y_t^{'}$): $y_t^{'} = w_t(y_t - \mu)$. At last, the corrected temporal derivative was integrated in order to obtain the corrected signal ($x_t^{'}$):

$$x_t^{'} = \sum_{i=1}^{N} (y_t^{'}) \tag{4}$$

After the motion artifacts are corrected, the channels were baselined. At the beginning of every round, a 30-s resting period was recorded to serve as a baseline. The average of the last 15 s of the resting period was used to subtract from all datapoints of the signal. The baselining process was done for each channel separately. Two channels, namely channels 4 and 11 (related to the motor cortex and the left PFC, respectively), were removed due to their ultra-low fNIRS activities possibly caused by bad optode connections.

Despite applying motion artifact correction, it is still possible that artifact effect still remained in the form of unreliable fNIRS waveforms. In normal situations where artifact is absent, the direction of concentration changes of the chromophores oxygenated hemoglobin (O2Hb signal) is opposite to that of deoxygenated hemoglobin (HHb signal), and therefore negative correlation of O2Hb and HHb can be expected. Nevertheless, motion artifacts in the signals can lead to concurrent change of both signals in the same direction, leading to more positive

[2]https://gopro.com/en/th/shop/hero7-black/tech-specs?pid=CHDHX-701-master

[3]https://store.netgate.com/Parani-SENA-Bluetooth-Adapter-UD100-G03-P1350.aspx

correlation. To measure the extent to which the fNIRS signals were affected by motion artifacts, we therefore calculated the correlation coefficient between both signals per trial and channel. It was suggested in the literature that large head movements can already increase the correlation coefficient to 0.1 (Cui et al., 2010) and jumping artifact can enhance the coefficient to 0.4 (Lee et al., 2018). As the movement during the experiment is comparable to jumping, we opted to use a higher threshold and the validity of the threshold has been confirmed by our empirical study on the threshold effect. Therefore, data for certain channels of certain trials were removed if the correlation coefficient between the O2Hb signal and the HHb signal was larger than a threshold of $\rho = 0.4$. The noisy data removal was done at individual trial-channel level, i.e., only noisy channel data was removed per trial rather than discarding the whole trial data. Consequently, ∼41% of the all trial-channel data remained.

2.4.2. Feature Extraction

Afterwards, we extracted features from the cleaned fNIRS signals in valid trials and channels.

1. Motor cortex activation; three mean features were obtained from three remained channels (channels 1, 2, and 3) related to this cortex.
2. Left temporal cortex activation; four mean features were obtained from three remained channels (channels 13, 14, 15, and 16) related to this cortex.
3. Averaged PFC activation; as there was no channel lying exactly in the middle between right and left hemispheres, we calculated a feature to represent PFC activation by averaging fNIRS signal from one representative channel in the left hemisphere and one from the right hemisphere. After bad-channel removal, there were three channels relevant to the left PFC (channels 9, 10, and 12) and four channels relevant to the right PFC (channels 5, 6, 7, and 8), generating 12 possible left-right combinations. We derived all 12 mean features from all channel-pairs as the features.
4. PFC asymmetry; similar to the averaged PFC feature, we calculated hemispheric asymmetry from all 12 combinations of left and right PFC channels by subtracting a left PFC channel from a right PFC channel and then calculated the mean of the result as a feature (Hatfield and Kerick, 2007; Silveira et al., 2019). Therefore, a positive value corresponds to a higher right PFC activation relative to left PFC activation, on average.
5. Connectivity between DLPFC and motor cortex; we calculated the connectivity index by following the method of Nguyen et al. (2018). First, Pearson correlation coefficients (ρ) between two fNIRS channels were calculated per trial by:

$$\rho_{x,y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}, \tag{5}$$

   where X and Y denote channel data, σ_X and σ_Y refer to the standard deviation of channel X and Y, respectively, and $cov(X,Y)$ refers to the covariance between the two channels. In order to convert the sampling distribution of the Pearson correlation coefficients into the normal distribution, the obtained ρ values were transformed to z values using the Fischer z-transformation:

$$z = \frac{1}{2} ln(\frac{1+\rho}{1-\rho}). \tag{6}$$

   The connectivity index was computed for all combinations of motor cortex channels (channels 1, 2, and 3) and DLPFC channels (channels 17 and 18). The number of significant connections was then determined by counting the number of connections that had an absolute z-value greater than our pre-defined threshold of 0.6, which is corresponding to a correlation of $\rho \approx 0.54$ and motivated by the results of Nguyen et al. (2018). Apart from the feature derived from counting the number of significant connections, we also calculated the mean of absolute z-values of all corresponding connections as another feature.

Negative feature values correspond to the lower feature values during the *task* 5-s waiting time before task execution compared to feature values during the *resting* period, and positive values mean vice versa. Outliers, which were defined as the values that deviated from the corresponding means across all participants for more than three standard deviations, were then removed in each feature.

2.4.3. Statistical Analysis

In order to test the hypotheses with the obtained features, permutation statistical tests were used, as they made no assumption on the distribution of data. A total of 100,000 permutations were used, suggesting that the smallest possible p-value is 10^{-5}. The analysis was performed on the extracted features under three different studies, each of which compared two different conditions, that help answer our research questions. The hypotheses were made by following previous findings in the literature.

1. Experienced vs. Inexperienced players; it was hypothesized that motor cortex activation (three channels) is higher (Wolf et al., 2014) and left temporal cortex activation (four channels) is lower (Hatfield et al., 1982; Haufler et al., 2000; Wolf et al., 2015) in experienced players compared to inexperienced players. In total, seven statistical tests were made for the hypotheses in these two features.
2. Anxious vs. Non-Anxious players; it was hypothesized that motor cortex activation (three channels) is lower (Lee and Grafton, 2015), while the averaged PFC activation (12 channel-pairs) (Korb, 2010; Schweizer et al., 2013; Nosrati et al., 2016), PFC asymmetry (12 channel-pairs) (Hatfield and Kerick, 2007), and the connection between DLPFC and motor cortex (two indices) (Yoon et al., 2006; Clapp et al., 2009; Lee and Grafton, 2015) are higher in anxious players compared to non-anxious players. As the left temporal cortex was found related to the suppression of automated skills, which are possessed only by an experienced player, different hypotheses were made for experienced and inexperienced players. It was hypothesized that experienced players have higher left temporal cortex activation (four channels) when

being anxious compared to non-anxious experienced players as the automatic skill suppression does not function properly when being anxious (Zhu et al., 2011; Wolf et al., 2015). In contrast, the opposite hypotheses were made for the inexperienced players. In total, 3 + 12 + 12 + 2 + 4 + 4 = 37 statistical tests were made in this study of anxiety.

3. Scored vs. Missed penalties; the number of statistical tests was identical to the study of anxiety. In particular, it was hypothesized that motor cortex activation (three channels) is higher (Lee and Grafton, 2015), while the averaged PFC activation (12 channel-pairs) (Korb, 2010; Schweizer et al., 2013; Nosrati et al., 2016) and PFC asymmetry (12 channel-pairs) (Meyer et al., 2015; Silveira et al., 2019) are lower, and the connection between DLPFC and motor cortex (two indices) are higher (Yoon et al., 2006; Clapp et al., 2009; Lee and Grafton, 2015) when scoring the penalties, compared to when missing penalties. Again, the analysis of the left temporal cortex was done separately by experienced and inexperienced players; experienced players were hypothesized to exhibit lower left temporal cortex activation (four channels) when scoring (Wolf et al., 2015), and the opposite hypotheses were made for inexperienced players.

The total 7 + 37 + 37 = 81 statistical tests are also summarized in **Table 2** that enumerates all features and all studies in this research. A multiple-testing correction was done using a false discovery rate (FDR) test as correction procedure (Singh and Dan, 2006), with significance threshold $Q = 0.05$ and the number of statistical tests $m = 81$. It is noteworthy that the connectivity analysis between DLPFC and motor cortex was done only on data from the last round of the penalty kick, which should involve the highest level of pressure, as the connectivity indices were found to be related to choking under pressure.

2.4.4. Classifying Brain Data

In order to assess how well fNIRS data can be used to distinguish the different levels of experience, anxiety, and success in penalty shooting, classification was done separately in each study. In each classification, a single type of feature, except connectivity indices, was used in order to allow an investigation on which feature is the most powerful for distinguishing two classes. In addition to the mean feature as used in the statistical analysis in the previous section, we also calculated the standard deviation, the minimum value, and the maximum value as additional features for each trial. Support Vector Machines (SVMs) with linear kernels were trained and tested on the feature data, where 80% of total data were randomly selected as training data and the rest 20% were used as test data. The classification was implemented using *scikit-learn*[4] package of Python and evaluated by the accuracy and area under the receiver operating curve (ROC) between true-positive rate and false-positive rate. As random shuffling involved with training and testing, the classification was performed five times and the grand average and standard deviation of the accuracy were reported.

[4]https://scikit-learn.org/stable/

3. RESULTS

3.1. Behavioral Results

Table 1 shows the performance of the players as the percentage of scored penalties in each round, duration for which the players were looking at the goalkeeper, and ratings of satisfaction and motivation to score at the end of each round. Wilcoxon's Rank Sum statistical tests with Bonferroni correction were performed on the comparison between experienced and inexperience players. In addition, Kruskal Wallis one-way analysis of variance with Bonferroni correction was used to compare performances and scores between round 1, 2, and 3. The performance scores indicate that inexperienced players performed the worst in the last round, whereas experienced players had a similar performance in the second and the last round. Overall, experienced players performed better than inexperienced players. The exception is the first round, as experienced players scored less in this round. Inexperienced players took considerably more risks in the later rounds. **Figure 3** shows the placement (shot-accuracy) of each penalty, demonstrating that during the last round more penalties were shot over or wide by inexperienced players. On average, inexperienced players shot their penalties higher and wider per round. Interestingly, this is not the case for experienced players. Although they also shot their penalties higher on average, the horizontal placement did not change between the rounds. Furthermore, the shot power for both experienced and inexperienced players increased in the later rounds. A significantly poorer performance from inexperienced players was notable when comparing between the second and the last rounds (see **Table 1**). For experienced players, this decrease in performance was not apparent. In fact, they performed slightly better in the last round as compared to the second round although the difference was not significant.

Inexperienced players looked significantly longer at the goalkeeper during the last round, but this is not the case for experienced players that looked slightly shorter during the last round. The goalkeeper was a larger distracting factor for inexperienced players in the last round. The aim was to distract the player in the last round and this tactic seems to have been successful for inexperienced players. Again, this does not seem the case for experienced players, as they fixated for a shorter period at the goalkeeper during the last round. This can be explained by the fact that experienced players are more familiar with these distracting methods of a goalkeeper and therefore know how to keep their concentration under these circumstances.

After each round was explained, the players were briefly asked how confident they were and how many goals they thought they would score. We observed that players were less confident in the last round, as the expectation on the number of goals to be scored became lower in the later rounds, especially among the inexperienced group.

Also, the results of the SAS questionnaire show that the pressure was highest in the last round (see **Table 1**). Apart from the total anxiety score that is reported in the table, we also found that its compositing worry score and somatic anxiety score

TABLE 1 | Behavioral results showing the percentage of the penalties that were scored for both experienced (Exp) and inexperienced (Inexp) players in each round, the averaged duration (out of the 5-s waiting period) that the players were looking at the goalkeeper, the average and standard deviation of the reported SAS scores enumerated by total (T), worry (T), disruption (D), and somatic (S) scores, satisfaction ratings and motivation ratings at the end of each round; * indicates significant difference between Exp and Inexp at $p < 0.01$ (corrected by Bonferroni correction); a, b, and c indicate significant differences at $p < 0.05$ (corrected by Bonferroni correction) between rounds 1–2, 1–3, and 2–3, respectively.

Round	Scored penalties (%)		Keeper-looking duration(s)		SAS				Satisfaction	Motivation
	Non-Exp	Exp	Non-Exp	Exp	T	W	D	S		
R1	98.3^{ab}	88.0^{ab}	–	–	29.86 ± 8.03	11.09 ± 3.94	6.91 ± 1.80	11.86 ± 4.37	3.45	3.59^{b}
R2	25.0^{*a}	60.0^{*a}	1.71^{c}	2.82	32.09 ± 8.33	12.86 ± 3.91	6.50 ± 1.50	12.73 ± 4.60	3.45	4.27
R3	18.3^{*b}	62.0^{*b}	3.41^{c}	2.41	34.32 ± 9.30	13.91 ± 5.06	6.91 ± 1.72	13.50 ± 4.75	2.82	4.59^{b}

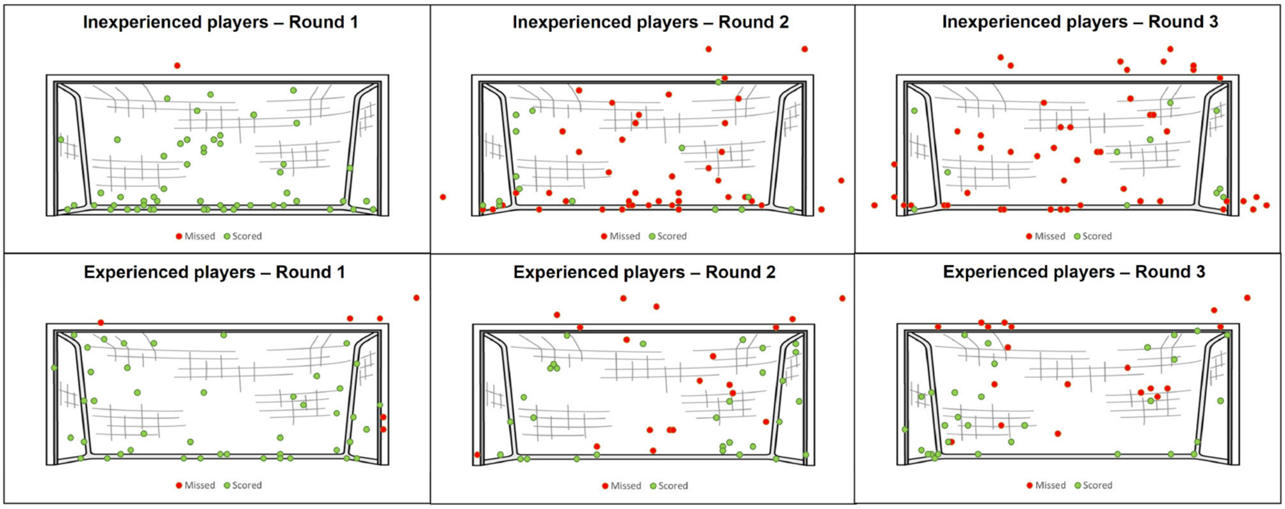


FIGURE 3 | Overview of the placement of penalty-kick execution for experienced and inexperienced players in each round. Red dots represent missed penalties and green dots represent scored penalties.

were lowest in the first round and highest in the last round. The concentration disruption score did not change between the rounds. For 12 out of the 22 participants, the total anxiety score increased per round. During the first round, six participants were considered to be at least *somewhat* anxious. This is determined by satisfying one of four following conditions; total score was above or equal to 42; worry score was above or equal to 14; disruption score was above or equal to 10; somatic score was above or equal to 18. During the second round, this number increased to nine participants, and during the last round, this number increased to twelve participants. In total, during 27 rounds (equivalent to 135 trials) out of the entire 66 rounds (namely 330 trials), the players reported to be at least *somewhat* anxious, which corresponds to 40.9% of the trials. Anxious players also missed more penalties (around 58%) than non-anxious players (around 31%). Furthermore, **Table 1** shows that the motivation rating was higher in the second round and significantly greater in the last round. Meanwhile, the satisfaction rating was lowest in the last round. As anxious players performed worse (more misses), it can be assumed that the results of the SAS questionnaire are trustworthy. Anxious players were more likely to miss (Wilson et al., 2009).

Figure 4 shows correlation coefficients between each behavioral resultant score considering all players, only experienced players, and only inexperienced players. Correlations were computed by Spearman's rank correlation method where significant results at p <0.05 (corrected by Bonferroni correction) are surrounded by red borders. Apart from trivial correlation within SAS scores, it can be observed that percentage of goal scoring is negatively correlated with SAS total and SAS worry scores when taking scores from all players into account. This indicates that anxiety can adversely affect performance in general. Considering experienced players, the scoring percentage is correlated with satisfaction but negatively correlated with SAS worry scores. It suggests that successful performance can strongly lead to satisfaction with the penalty kick in this group, while anxiety can hinder the performance. In contrast, SAS worry scores in inexperienced players are correlated with motivation. It can be inferred that willingness to score can enhance worrisome in novice players or vice versa.

However, fNIRS data of one participant were discarded from the subsequent analysis due to technical failure in the recording. Statistical analysis and classification were done on the data from the remaining 21 participants.

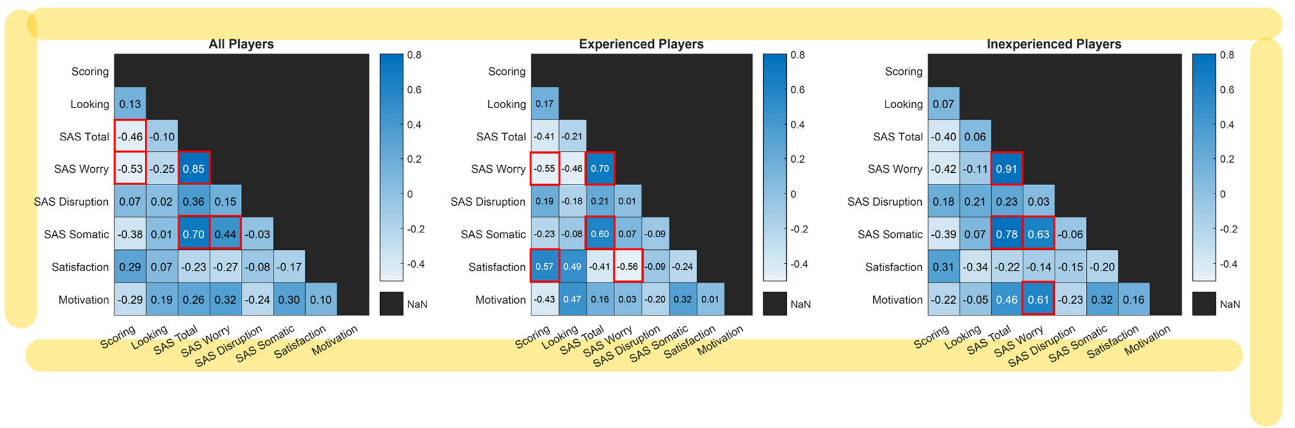


FIGURE 4 | Correlation coefficients between each behavioral resultant scores in all players, experienced players, and inexperienced players. Significant correlations at $p < 0.05$ (corrected by Bonferroni correction) are with red borders.

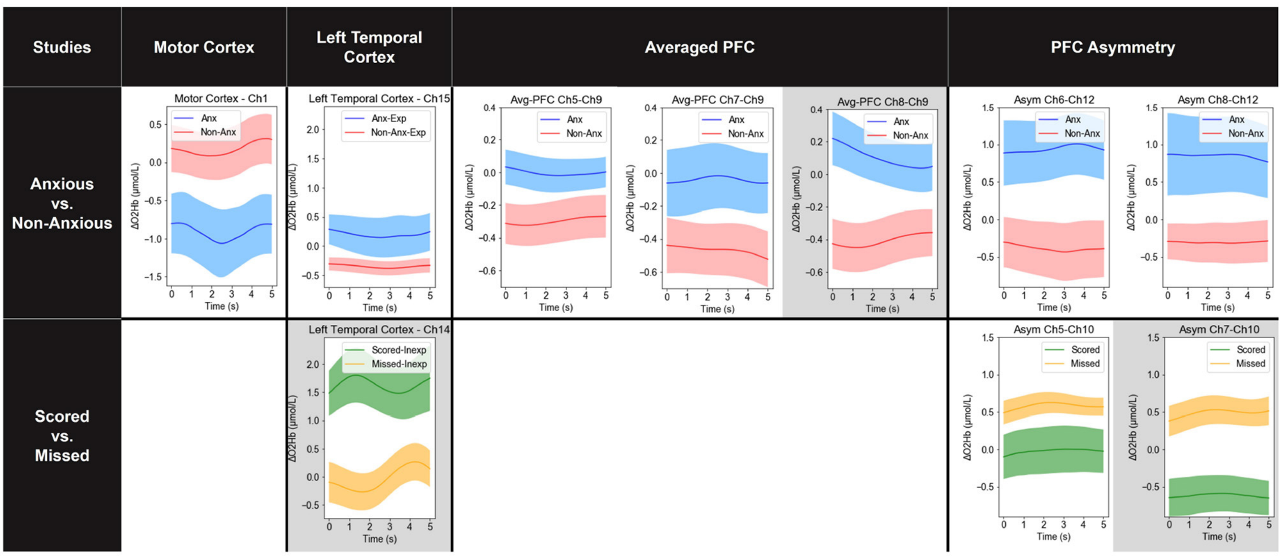


FIGURE 5 | Averaged features across all trials from all participants where the difference between two conditions in a specific study, regarding experience, anxiety, and success/failure, are significant at $p < 0.05$ (uncorrected). Features are shown in the variation over a 5-s waiting period. Note that the studies of anxiety and success/failure in the left temporal cortex were done separately by groups of experience [experienced (Exp) and inexperienced (Inexp) players]. Results that remain statistically significant at $p < 0.05$ after FDR correction are with a gray background.

3.2. Statistical Analysis

As O2Hb concentration is directly related to the activation of a brain area, we only focus on O2Hb concentration contrast with the 15-s baseline period preceding each round and discard HHb data from our analysis. **Figure 5** shows all testings where the feature values for the two conditions differ at the significance level p <0.05 in a particular study. Tests that remained providing significant results with p <0.05 after FDR correction were labeled with a gray background. A blank cell means no significant results were found in the test. The mean and standard deviation of the mean values for each test were summarized in **Table 2**.

The trials of all rounds were included for the comparison between experienced and inexperienced players but significant results were not found. Regarding anxiety, the results in **Figure 5** show that in the motor cortex, the difference between anxious and non-anxious players was the largest for channel 1, where the cortex was clearly less activated for anxious players. The averaged PFC activation was greatly higher in channel-pairs 5–9, 7–9, and 8–9 in anxious players. Whilst being anxious, a right PFC activation was found higher compared to left PFC activation as evidenced by more PFC asymmetry in channel-pairs 6–12 and 8–12. For the left temporal cortex, experienced and inexperienced players were analyzed separately, as the hypotheses suggest that there could be a difference between the two. Anxious experienced players showed a clear higher activation in channel 15 compared to those who were not anxious, while no significant results were found from inexperienced players. Similarly, DLPFC-motor cortex connectivity analysis on the last-round data did not indicate any significant difference between anxious and non-anxious players.

The contrast between a successful performance (scoring) and a failed performance (missing) can be reflected mostly by the PFC asymmetry, especially in channel-pairs 5–10 and 7–10 (see **Figure 5**). Again, for the left temporal cortex, the analyses for experienced and inexperienced players were analyzed separately. The results suggest that inexperienced players showed an increased left temporal cortex activation when scoring. Similarly, DLPFC-motor cortex connectivity analysis in the last-round penalty kick (with highest pressure) did not indicate any significant difference between scoring and missing.

3.3. Classifier Results

Table 3 shows the classification results in each study using different features. The best result was obtained by using the motor cortex feature to distinguish between experienced and inexperienced players, achieving 66.7% of accuracy and 0.6806 area under ROC. In general, anxiety and non-anxiety were classified most correctly by motor cortex data. The informative feature to classify scored penalties against missed penalties is the averaged PFC activation feature based on accuracy and the left temporal cortex feature based on area under ROC.

Considering the supposed chance-level of 50%, all classification results were above the chance level but with a small margin. This led us to the analysis of data distribution and its impacts on the classification performance. Specifically, principal component analysis (PCA) was applied to some features in each classification, i.e., motor cortex data for experienced vs. inexperienced players classification, averaged PFC for anxious vs. non-anxious players classification, and PFC asymmetry for scored vs. missed penalties classification. Then, we visualized the distribution of data that were projected into the first and second principal components (PCs) as shown in **Figure 6**. Apparently, datapoints of both classes were clustered together, instead of nicely spreading into different locations in dimensional space. Therefore, it is difficult for a linear classifier to achieve high performance in the classification task.

4. DISCUSSION

In the present study, we demonstrated the feasibility to explore brain activity *in the field* prior to executing a penalty kick, which is a strenuous physical activity that has been challenging neuroimaging research (Carius et al., 2020). Our results show neurological evidence in fNIRS signals that are related to the level of experience in soccer, anxiety before task execution, and scoring success/failure.

4.1. Success in Pressure Induction

The poorer scoring performance in the second and last rounds, over-bar, and wider shots for missed penalties over round, and the increment of shot powers per round can be observed from the results. These all indicate that the players took more risks in the last round. This can be the indication that the pressure was successfully induced. Especially, inexperienced players seem to have experienced a higher level of pressure. There must have been other factors involved in this phenomenon. In the last round, the players were namely competing for a prize and in order to win this prize, they had to not only score the most goals but also to create the best-quality goals. We speculated that the incentive could have influenced them to take more risks in the last round, e.g., by trying to shoot the ball in the top corner.

Significantly poorer performance from inexperienced players could be an indication of heightened mental pressure. It could also be explained by the fact that the players had already taken five penalties against the goalkeeper. Based on these five penalties, the goalkeeper could potentially already know what the shooting technique and favorite corner of the player would be. In contrast, the comparable performance of experienced players between the second and the last round can be observed. Some experienced players verbally reported that they needed a certain *eustress* in order to well perform, which could be an explanation of the slight increase in performance.

At the end of the experiments, the players were also asked how much pressure they experienced in each round and the majority indicated that they experienced the most pressure in the last round. Overall, it can be concluded that pressure was successfully induced as reflected by most indicators. The distribution between anxious and non-anxious players is also nicely balanced (41–59%).

4.2. Results in Line With Neural Efficiency Theory

Focusing on anxiety analysis, it is implied that our results are mostly in line with the neural efficiency theory. When being anxious, the motor cortex (task-relevant area) was activated significantly less in one channel. The activation of task-irrelevant areas of the brain was more common when being anxious. This was most prominently observable in the PFC, as a significant increase in averaged PFC activation in three channel-pairs was related to being more anxious. These results are in line with the previous works (Korb, 2010; Schweizer et al., 2013; Nosrati et al., 2016) that reported the association between overactivation of PFC and choking under pressure. According to Korb (2010), this overactivation would cause a distraction, decreasing one's focus on the task. The results of the present study agree with such theory, as an increase in PFC activation was paired with a decrease in motor cortex activation when being anxious. The long-term thinking element of the PFC could be the source of this distraction, as players might think about the consequences of missing or scoring the penalty (Korb, 2010). Besides this increase in averaged PFC activation, the anxiety level of the player was also notable in the difference between left and right PFC activation. For two channel-pairs, the right PFC was more

TABLE 2 | Statistics [mean and standard deviation (Std)] and results from the statistical test (*p*-values) comparing features from two conditions in a specific study regarding experience, anxiety, and success/failure.

Studies	Features	Channels		Participants			Null hypothesis of tests	Condition (1)		Condition (2)		*p*-Value
		Chan1	Chan2	All	Exp	In-exp		Mean	Std	Mean	Std	
	MC	1		•			(1) >(2)	0.129	0.095	−0.617	0.114	0.071
	MC	2		•			(1) >(2)	0.067	0.101	0.166	0.054	0.621
Experienced (1)	MC	3		•			(1) >(2)	−0.442	0.045	0.577	0.083	0.997
vs.	LTC	13		•			(1) <(2)	−0.162	0.035	−0.102	0.006	0.429
Inexperienced (2)	LTC	14		•			(1) <(2)	0.296	0.056	0.633	0.101	0.190
	LTC	15		•			(1) <(2)	−0.268	0.026	0.029	0.028	0.053
	LTC	16		•			(1) <(2)	0.208	0.041	0.116	0.033	0.626
	MC	**1**		**•**			**(1) <(2)**	**−0.911**	**0.093**	**0.173**	**0.078**	**0.015**
	MC	2		•			(1) <(2)	−0.126	0.116	0.283	0.038	0.103
	MC	3		•			(1) <(2)	0.171	0.094	−0.002	0.036	0.675
	LTC	13			•		(1) >(2)	−0.037	0.023	−0.223	0.045	0.294
	LTC	14			•		(1) >(2)	0.086	0.126	0.341	0.042	0.642
	LTC	**15**			**•**		**(1) >(2)**	**0.194**	**0.039**	**−0.348**	**0.024**	**0.046**
	LTC	16			•		(1) >(2)	0.266	0.070	0.193	0.034	0.444
	LTC	13				•	(1) <(2)	−0.173	0.024	0.007	0.043	0.372
	LTC	14				•	(1) <(2)	0.610	0.133	0.665	0.098	0.465
	LTC	15				•	(1) <(2)	−0.044	0.027	0.211	0.051	0.211
	LTC	16				•	(1) <(2)	−0.045	0.071	0.483	0.146	0.138
	Asym	5	9	•			(1) >(2)	0.009	0.030	0.597	0.041	0.958
	Asym	5	10	•			(1) >(2)	0.503	0.059	0.168	0.025	0.182
	Asym	5	12	•			(1) >(2)	0.393	0.020	0.024	0.050	0.143
	Asym	6	9	•			(1) >(2)	0.387	0.035	0.051	0.039	0.239
	Asym	6	10	•			(1) >(2)	0.265	0.072	0.058	0.029	0.344
	Asym	**6**	**12**	**•**			**(1) >(2)**	**0.941**	**0.041**	**−0.387**	**0.037**	**0.010**
Anxious (1)	Asym	7	9	•			(1) >(2)	−0.066	0.051	−0.231	0.028	0.374
vs.	Asym	7	10	•			(1) >(2)	0.135	0.079	−0.281	0.023	0.121
Non-Anxious (2)	Asym	7	12	•			(1) >(2)	0.073	0.078	−0.186	0.021	0.246
	Asym	8	9	•			(1) >(2)	−0.266	0.018	0.052	0.010	0.877
	Asym	8	10	•			(1) >(2)	0.108	0.057	0.103	0.044	0.503
	Asym	**8**	**12**	**•**			**(1) >(2)**	**0.848**	**0.026**	**−0.307**	**0.008**	**0.016**
	Avg-PFC	**5**	**9**	**•**			**(1) >(2)**	**−0.005**	**0.015**	**−0.299**	**0.020**	**0.041**
	Avg-PFC	5	10	•			(1) >(2)	−0.252	0.029	−0.084	0.012	0.817
	Avg-PFC	5	12	•			(1) >(2)	−0.197	0.010	−0.012	0.025	0.857
	Avg-PFC	6	9	•			(1) >(2)	−0.083	0.054	−0.033	0.015	0.576
	Avg-PFC	6	10	•			(1) >(2)	−0.256	0.077	−0.234	0.042	0.519
	Avg-PFC	6	12	•			(1) >(2)	−0.451	0.058	−0.164	0.058	0.797
	Avg-PFC	**7**	**9**	**•**			**(1) >(2)**	**−0.042**	**0.016**	**−0.468**	**0.021**	**0.048**
	Avg-PFC	7	10	•			(1) >(2)	−0.255	0.054	−0.232	0.023	0.531
	Avg-PFC	7	12	•			(1) >(2)	−0.679	0.040	−0.294	0.018	0.927
	Avg-PFC	**8**	**9**	**•**			**(1) >(2)**	**0.101**	**0.059**	**−0.410**	**0.034**	**0.011***
	Avg-PFC	8	10	•			(1) >(2)	−0.162	0.038	−0.483	0.047	0.062
	Avg-PFC	8	12	•			(1) >(2)	−0.357	0.015	−0.238	0.027	0.728
	#Con	–	–	•			(1) >(2)	0.544	0.287	0.487	0.288	0.255
	z-val	–	–	•			(1) >(2)	0.935	0.607	0.811	0.430	0.217
	MC	1		•			(1) >(2)	−0.197	0.062	−0.462	0.116	0.308
	MC	2		•			(1) >(2)	0.254	0.093	−0.126	0.043	0.122
	MC	3		•			(1) >(2)	0.125	0.047	−0.005	0.101	0.374
	LTC	13			•		(1) <(2)	0.007	0.039	−0.510	0.034	0.937

(Continued)

TABLE 2 | Continued

Studies	Features	Channels		Participants			Null hypothesis of tests	Condition (1)		Condition (2)		*p*-Value
		Chan1	Chan2	All	Exp	In-exp		Mean	Std	Mean	Std	
	LTC	14			•		(1) <(2)	0.599	0.068	−0.418	0.041	0.989
	LTC	15			•		(1) <(2)	−0.379	0.025	−0.021	0.033	0.074
	LTC	16			•		(1) <(2)	0.425	0.063	−0.388	0.045	0.970
	LTC	13				•	(1) >(2)	−0.605	0.037	0.315	0.036	0.957
	LTC	**14**				**•**	**(1) >(2)**	**1.634**	**0.108**	**−0.035**	**0.196**	**0.001***
	LTC	15				•	(1) >(2)	0.068	0.077	0.007	0.011	0.416
	LTC	16				•	(1) >(2)	0.288	0.218	−0.004	0.157	0.260
	Asym	5	9	•			(1) <(2)	0.189	0.019	0.466	0.032	0.209
	Asym	**5**	**10**	**•**			**(1) <(2)**	**−0.023**	**0.026**	**0.582**	**0.035**	**0.039**
	Asym	5	12	•			(1) <(2)	0.142	0.045	0.178	0.025	0.461
	Asym	6	9	•			(1) <(2)	0.007	0.030	0.506	0.047	0.152
	Asym	6	10	•			(1) <(2)	−0.048	0.012	0.416	0.067	0.178
	Asym	6	12	•			(1) <(2)	0.255	0.035	0.042	0.075	0.631
	Asym	7	9	•			(1) <(2)	−0.403	0.015	0.148	0.032	0.136
Scored (1) vs. Missed (2)	**Asym**	**7**	**10**	**•**			**(1) <(2)**	**−0.616**	**0.020**	**0.488**	**0.041**	**0.0003***
	Asym	7	12	•			(1) <(2)	−0.043	0.037	−0.203	0.046	0.661
	Asym	8	9	•			(1) <(2)	−0.111	0.012	−0.047	0.015	0.407
	Asym	8	10	•			(1) <(2)	0.003	0.021	0.214	0.062	0.292
	Asym	8	12	•			(1) <(2)	−0.148	0.048	0.532	0.053	0.112
	Avg-PFC	5	9	•			(1) <(2)	−0.094	0.010	−0.233	0.016	0.790
	Avg-PFC	5	10	•			(1) <(2)	0.011	0.013	−0.291	0.018	0.962
	Avg-PFC	5	12	•			(1) <(2)	−0.071	0.023	−0.089	0.013	0.541
	Avg-PFC	6	9	•			(1) <(2)	0.172	0.035	−0.449	0.108	0.991
	Avg-PFC	6	10	•			(1) <(2)	−0.085	0.038	−0.479	0.022	0.856
	Avg-PFC	6	12	•			(1) <(2)	−0.218	0.043	−0.475	0.089	0.748
	Avg-PFC	7	9	•			(1) <(2)	−0.233	0.040	−0.333	0.030	0.649
	Avg-PFC	7	10	•			(1) <(2)	−0.190	0.025	−0.305	0.021	0.669
	Avg-PFC	7	12	•			(1) <(2)	−0.393	0.016	−0.481	0.027	0.631
	Avg-PFC	8	9	•			(1) <(2)	−0.186	0.027	−0.205	0.015	0.534
	Avg-PFC	8	10	•			(1) <(2)	−0.375	0.020	−0.346	0.031	0.443
	Avg-PFC	8	12	•			(1) <(2)	−0.343	0.014	−0.200	0.040	0.232
	#Con	–	–	•			(1) <(2)	0.668	0.330	0.505	0.298	0.052
	z-val	–	–	•			(1) <(2)	1.061	0.635	0.831	0.472	0.098

Features were extracted from channels related to the motor cortex (MC) and the left temporal cortex (LTC), from channel-pairs demonstrating PFC asymmetry (Asym) and averaged PFC (Avg-PFC), and from DLPFC-MC connectivity showing the averaged number of significant connections (♯Con) and averaged z-scored connectivity index (z-val). Note that the studies of anxiety and success/failure in LTC were done separately by groups of experience (experienced (Exp) and inexperienced (Inexp) players). Results that are statistically significant at $p < 0.05$ (uncorrected) are in bold, and those that are significant after FDR correction are marked with asterisks.

activated compared to the left PFC for anxious players; this is in line with the previous findings (Hatfield and Kerick, 2007; Meyer et al., 2015; Silveira et al., 2019) that imbalanced PFC activation (caused by a stronger right PFC activation) leads to choking under pressure. For experienced players, increased activation of the left temporal cortex was linked with being anxious; this is in line with previous evidence (Zhu et al., 2010; Wolf et al., 2015) that the left temporal cortex's relationship to self-instruction and self-reflecting can cause a distraction for experienced players. Experienced players should trust on their automated skills and therefore do need to suppress self-instruction and self-reflection processes, which are essential skills in the early stages of learning a motor skill (Wolf et al., 2015). By activating the left temporal cortex more, experienced players neglect their automated skills and start to *overthink* the situation. This increase can be seen as a distracting factor.

An increase in left temporal cortex activation was expected to be related to neglecting automated skills and therefore to poorer performance (i.e., missing penalties) among experienced players. For inexperienced players, the opposite trend was expected and indeed observed in one channel in our results. For PFC asymmetry, however, the results were greatly in line with the literature (Hatfield and Kerick, 2007; Meyer et al., 2015; Silveira et al., 2019). For channel-pairs 5–10 and 7–10, the right

TABLE 3 | Mean accuracy, standard deviation (in brackets), and area under the ROC (in italic style) of SVM-based classification from five runs using a different type of features.

Features	Exp/Inexp	Anx/Non-Anx	Scored/Missed
Motor cortex	66.7% (±5.1)	61.5% (±10.3)	50.8% (±4.2)
	0.6806	*0.6048*	*0.4618*
Left temporal cortex	62.4% (±7.8)	58.8% (±9.1)	56.8% (±5.6)
	0.6202	*0.5302*	*0.5458*
Averaged PFC		54.6% (±7.8)	60.5% (±9.0)
		0.4980	*0.4916*
PFC asymmetry		59.2% (±3.6)	59.5% (±3.9)
		0.5094	*0.4988*

The classification was performed per study; namely experienced (Exp) vs. inexperienced (Inexp) players; anxious (Anx) vs. non-anxious (Non-Anx) players; and scored vs. missed penalties.

PFC was more activated, as compared to the left PFC, when missing a penalty.

It is noteworthy that when using an FDR-correction, with $Q = 0.05$ and $m = 81$, 3 out of the 10 significant results remain significant. The FDR correction is more often applied to channel-wise fNIRS analyses, similar to this study (Singh and Dan, 2006). These FDR-corrected results imply that most of the significant results could be a coincidence. The only significant results that remain after the correction are: the left PFC is more activated than the right PFC when scoring a penalty, anxious players show a higher averaged PFC activation, and inexperienced players show an increased left temporal cortex activation when scoring a penalty. Although most results are not significant after FDR-correction, these results are still in line with previous findings in the literature. Therefore, although no direct conclusions can be drawn for the results of this study alone, the results can still be seen as a support of the theory in the literature.

4.3. Limitations of the Study

The greatest challenge of this study is similar to other *in situ* studies—motion artifacts, despite the fact that the fNIRS technology is less susceptible to motions artifacts and electrical noise. Although the participants were instructed to minimize their movement during the 5-s waiting period before whistle signal to start executing the kick, the intensive eagerness to perform the task led to undesirable tiny movements in certain participants, resulting in the loss of 60% of total data. A solution would be to prolong the waiting period to 10 s, which might help decrease the probability of motion artifacts in the signal. Also, prolonging the waiting period would enable an alternative baselining method to utilize the early period of the trial as a baseline and allow the comparison with baselining by 30-s resting period in the current study.

Within the current study, scoring a penalty was seen as a successful performance, and missing a penalty is considered as a failed performance. However, this may not be the best measurement to use for this comparison. Scoring a penalty does not necessarily indicate that the penalty was taken well. For example, the goalkeeper can make a mistake, meaning that a badly taken penalty can still be a goal. In contrast, missing a penalty does not necessarily mean that a penalty was taken badly, as a goalkeeper can still save a penalty by correctly guessing the direction. It would therefore be recommended to, instead, look at the quality of the penalty. This can be done by, for example, looking at the shot-placement and shot-power.

Although mental pressure was successfully induced during this experiment, the levels of pressure were not the same as in professional soccer matches. The level of pressure during an important (professional) soccer match was not met and therefore it is uncertain if the pressure was sufficiently high to induce choking. A way to increase the level of pressure during an experiment is to recruit more spectators to witness penalty kicks.

4.4. Recommendations for Future Work

Future research should consider adding more trials per condition and prolonging the duration of each trial. This would allow performing reliable statistical analyses and calculating heart rate variability, which can be captured from the embedded cardiac cycles in the fNIRS signals and was found to be a useful measure to detect stress and choking (Taelman et al., 2009). However, this has to be compromised with potential fatigue, which was reported by the participants as minimal because the task execution in this experiment lasted for about 25 min on average (std = 2.32 min). Besides, longer trials would enable the application of the sliding window technique, which was found to improve accuracy in detecting a mental state (mind-wandering) (Liu et al., 2020). Repeating the experiment with similar protocol to this study can also allow the comparison of classification methods.

The statistical analyses of the fNIRS data and the classification performances are merely based on the mean of O2Hb features. However, it is known from previous fNIRS studies investigating mental states that alternative features, such as amplitude, slope, standard deviation, kurtosis, skewness, and signal peaks can provide insights and be used as discriminative features for classifying mental states. It is anticipated that alternative features, such as the maximum signal value, the time to peak, and the signal slope have the potential to improve the classification results.

Other factors that can influence mental states and affect the results are also worth investigation in future works, such as the interaction between penalty takers and goalkeepers, weather condition, comfortability of the fNIRS headset, amateur vs. professional players, the noise-sensitivity of the methods (Veale, 2013; Molavi et al., 2014), and the inter-subject variability in pressure induction.

It is noteworthy that the goal of this study is not to find the best classification model but to examine to what extent a simple linear classifier with minimal parameter tuning can classify different levels of experience, anxiety, success in penalty shooting. In our case, SVMs with linear kernels were employed and achieved 66.7% of accuracy and 0.6806 area under ROC at maximum of classification task. Future works can further improve the performance of classification by applying sophisticated algorithms, therefore the results in this study can only serve as a baseline. As we encourage other researchers to test

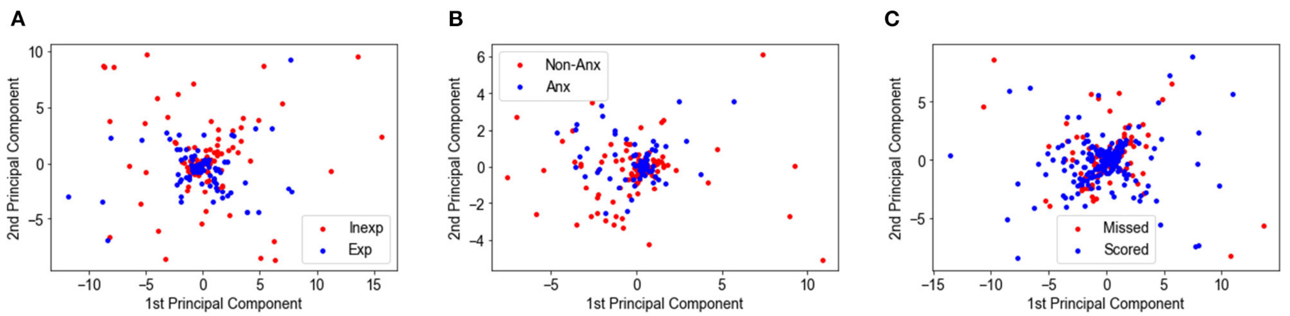


FIGURE 6 | Data distribution projected to first and second principal components for classifying: **(A)** experienced vs. inexperienced players using motor cortex data; **(B)** anxious vs. non-anxious players using averaged PFC data; **(C)** scored vs. missed penalties using the PFC asymmetry data.

other classification paradigms, we made the physiological data publicly available.

We believe that neurofeedback regarding neural efficiency can have implications not limited to the soccer domain but also in other professions and tasks where physical performance under pressure is essential.

5. CONCLUSION

In the present study, a penalty-kick experiment *in the field* was set-up, where pressure was successfully induced. Our results provide supportive evidence for the neural efficiency theory where the *correct* regions of the brain should be activated to successfully perform motor tasks under mental pressure. We demonstrated that brain activity associated with choking under pressure in a penalty kick situation can be reflected by *in-the-field* fNIRS measurement.

The results help answer our defined research questions. Regarding RQ1 that focused on performance, we related our findings with neural efficiency theory, demonstrating that the task-irrelevant PFC was related to missing penalties. This PFC activation showed itself in a higher right PFC activation compared to left PFC activation. The activation of the PFC can infer a distraction. This distraction is potentially caused by the long-term thinking ability of the PFC, as players might concern about the consequences of scoring or missing the penalty. However, we expected that connectivity between the motor cortex and the DLPFC during the last round of task execution (when mental pressure was highest) should provide insights on performing under high pressure, but no significant results were found. We therefore cannot answer this question. Similarly, we did not find significant difference in brain activity between experienced and inexperienced soccer players when taking a penalty kick to answer our RQ3.

We found that experienced players showed a higher left temporal cortex activation when being anxious, answering our RQ4 that focuses on anxious experienced players. As the left temporal cortex is related to self-instruction and self-reflection, this increased left temporal cortex activation indicates that experienced players *overthink* the situation and neglect their automated skills.

Focusing on our RQ5 related to anxious inexperienced players, no significant results were found. However, when discarding level of expertise, we found that the averaged PFC activation was also related to players with anxiety. Similarly, an increased right PFC activation, as compared to left PFC activation, was shown to be related to anxious players, irrespective of the level of expertise. Also, the motor cortex tends to have lower activation when being anxious regardless of the experience group.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: [https://doi.org/10.4121/14453556].

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente (reference number: RP 2020-118). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MS, NT, and MP conceived, planned, and designed the experiment. MS provided a critical review, carried out the experiment, data acquisition, analysis, processing interpretation, and wrote the full research report. NT wrote the manuscript with input from all the authors. MP supervised the research, provided the critical feedback, and proofread the manuscript. All authors discussed the results and contributed to the final manuscript.

FUNDING

This work was partially supported by the European Regional Development Fund's operationeel programma oost (OP-OOST

EFRO PROJ-00900) and by the Netherlands Organization for Scientific Research (NWA Startimpuls 400.17.602).

ACKNOWLEDGMENTS

The authors gratefully acknowledge soccer clubs vv Drienerlo and vv Den Ham for their help and support in the experiment, especially Silvie van de Boogaard, Vincent Delke, Allyne Groen, Patrick van Oerle, and Harry Vaassen. The authors also thank Job van Regteren, Nicole Rohring at University of Twente (UT) Sports Center, faculty members and staffs at the Faculty of Behavioural, Management and Social Sciences, UT, and Liucija Svinkunaite at Artinis for materials and facilities supports in this study. The authors also thank Sandra van Essen at Dutch overarching sports organization for discussion in experimental design and Dirk Heylen at Human-Media Interaction group, UT for advice and discussion.

REFERENCES

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

Al-shargie, F., Tang, T. B., and Kiguchi, M. (2016). "Mental stress grading based on fNIRS signals," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Orlando, FL: IEEE), 5140–5143. doi: 10.1109/EMBC.2016.7591884

Arrondel, L., Duhautois, R., and Laslier, J. F. (2019). Decision under psychological pressure: the shooter's anxiety at the penalty kick. *J. Econ. Psychol*. 70, 22–35. doi: 10.1016/j.joep.2018.10.008

Baumeister, R. F. (1984). Choking under pressure: self-consciousness and paradoxical effects of incentives on skillful performance. *J. Pers. Soc. Psychol*. 46, 610–620. doi: 10.1037/0022-3514.46.3.610

Carius, D., Hörnig, L., Ragert, P., and Kaminski, E. (2020). Characterizing cortical hemodynamic changes during climbing and its relation to climbing expertise. *Neurosci. Lett*. 715:134604. doi: 10.1016/j.neulet.2019.134604

Chiappori, P., Levitt, S., and Groseclose, T. (2002). Testing mixed-strategy equilibria when players are heterogeneous: the case of penalty kicks in soccer. *Am. Econ. Rev*. 92, 1138–1151. doi: 10.1257/00028280260344678

Chiarelli, A. M., Maclin, E. L., Fabiani, M., and Gratton, G. (2015). A kurtosis-based wavelet algorithm for motion artifact correction of fNIRS data. *Neuroimage* 112, 128–137. doi: 10.1016/j.neuroimage.2015.02.057

Clapp, W. C., Rubens, M. T., and Gazzaley, A. (2009). Mechanisms of working memory disruption by external interference. *Cereb. Cortex* 20, 859–872. doi: 10.1093/cercor/bhp150

Cohen-Zada, D., Krumer, A., Rosenboim, M., and Shapir, O. M. (2017). Choking under pressure and gender: evidence from professional tennis. *J. Econ. Psychol*. 61, 176–190. doi: 10.1016/j.joep.2017.04.005

Cui, X., Bray, S., and Reiss, A. L. (2010). Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics. *Neuroimage* 49, 3039–3046. doi: 10.1016/j.neuroimage.2009.11.050

Eberspächer, H., Gabler, H., and Hahn, E. (1990). *Praxis der Psychologie im Leistungssport*. Berlin: Philippka.

Fishburn, F. A., Ludlum, R. S., Vaidya, C. J., and Medvedev, A. V. (2019). Temporal derivative distribution repair (TDDR): a motion correction method for fnirs. *Neuroimage* 184, 171–179. doi: 10.1016/j.neuroimage.2018.09.025

Fryer, A. M., Tenenbaum, G., and Chow, G. M. (2018). Linking performance decline to choking: players' perceptions in basketball. *J. Sports Sci*. 36, 256–265. doi: 10.1080/02640414.2017.1298829

Furley, P., Noël, B., and Memmert, D. (2017). Attention towards the goalkeeper and distraction during penalty shootouts in association football: a retrospective analysis of penalty shootouts from 1984 to 2012. *J. Sports Sci*. 35, 873–879. doi: 10.1080/02640414.2016.1195912

Genakos, C., and Pagliero, M. (2011). Interim rank, risk taking and performance in dynamic tournaments. *J. Polit. Econ*. 120, 782–813. doi: 10.1086/668502

González-Díaz, J., and Palacios-Huerta, I. (2016). Cognitive performance in competitive environments: evidence from a natural experiment. *J. Publ. Econ*. 139, 40–52. doi: 10.1016/j.jpubeco.2016.05.001

Gröpel, P., and Mesagno, C. (2017). Choking interventions in sports: a systematic review. *Int. Rev. Sport Exerc. Psychol*. 12, 176–201. doi: 10.1080/1750984X.2017.1408134

Hatfield, B., and Kerick, S. (2007). "The psychology of superior sport performance: a cognitive and affective neuroscience perspective," in *Handbook of sport Psychology*, Vol. 3, eds G. Tenenbaum and R. C. Eklund (Hoboken, NJ: John Wiley & Sons Inc.), 84–109. doi: 10.1002/9781118270011.ch4

Hatfield, B. D., Landers, D. M., Ray, W. J., and Daniels, F. S. (1982). An electroencephalographic study of elite rifle shooters. *Am. Marksman* 7, 6–8.

Haufler, A. J., Spalding, T. W., Maria, D. L. S., and Hatfield, B. D. (2002). Erratum to "neuro-cognitive activity during a self-paced visuospatial task: comparative eeg profiles in marksmen and novice shooters". *Biol. Psychol*. 59, 87–88. doi: 10.1016/S0301-0511(01)00122-3

Haufler, A. J., Spalding, T. W., Maria, D. S., and Hatfield, B. D. (2000). Neuro-cognitive activity during a self-paced visuospatial task: comparative eeg profiles in marksmen and novice shooters. *Biol. Psychol*. 53, 131–160. doi: 10.1016/S0301-0511(00)00047-8

Hickman, D. C., and Metz, N. E. (2015). The impact of pressure on performance: evidence from the PGA tour. *J. Econ. Behav. Organ*. 116, 319–330. doi: 10.1016/j.jebo.2015.04.007

Hill, D. M., Hanton, S., Matthews, N., and Fleming, S. (2010). Choking in sport: a review. *Int. Rev. Sport Exerc. Psychol*. 3, 24–39. doi: 10.1080/17509840903301199

Ito, H., Yamauchi, H., Kaneko, H., Yoshikawa, T., Nomura, K., and Honjo, S. (2011). Prefrontal overactivation, autonomic arousal, and task performance under evaluative pressure: a near-infrared spectroscopy (NIRS) study. *Psychophysiology* 48, 1563–1571. doi: 10.1111/j.1469-8986.2011.01220.x

Jahani, S., Setarehdan, S. K., Boas, D. A., and Yücel, M. A. (2018). Motion artifact detection and correction in functional near-infrared spectroscopy: a new hybrid method based on spline interpolation method and Savitzky–Golay filtering. *Neurophotonics* 5, 1–11. doi: 10.1117/1.NPh.5.1.015003

Jordet, G., Hartman, E., Visscher, C., and Lemmink, K. A. P. M. (2007). Kicks from the penalty mark in soccer: the roles of stress, skill, and fatigue for kick outcomes. *J. Sports Sci*. 25, 121–129. doi: 10.1080/02640410600624020

Kamran, M. A., Naeem Mannan, M. M., and Jeong, M. Y. (2018). Initial-dip existence and estimation in relation to DPF and data drift. *Front. Neuroinform*. 12:96. doi: 10.3389/fninf.2018.00096

Kerick, S., Douglass, L., and Hatfield, B. (2004). Cerebral cortical adaptations associated with visuomotor practice. *Med. Sci. Sports Exerc*. 36, 118–129. doi: 10.1249/01.MSS.0000106176.31784.D4

Korb, A. (2010). *Penalty Kicks and the Prefrontal Cortex*. Available online at: https://www.psychologytoday.com/intl/blog/prefrontal-nudity/201007/penalty-kicks-and-the-prefrontal-cortex?amp

Kuriyama, T., Asona, M., Nambu, M., and Yoshida, M. (2015). "Prefrontal activity of goal keeping when penalty kick," in *World Congress on Medical Physics and Biomedical Engineering*, Vol. 51, ed D. A. Jaffray (Cham: Springer), 1155–1158. doi: 10.1007/978-3-319-19387-8_280

Le Fevre, M., Matheny, J., and Kolt, G. (2003). Eustress, distress, and interpretation in occupational stress. *J. Manag. Psychol*. 18, 726–744. doi: 10.1108/02683940310502412

Lee, G., Jin, S. H., Yang, S. T., An, J., and Abibulaev, B. (2018). "Cross-correlation between HBO and HBR as an effective feature of motion artifact in fNIRS signal," in *2018 6th International Conference on Brain-Computer Interface (BCI)* (Gangwon), 1–3. doi: 10.1109/IWW-BCI.2018.8311513

Lee, T. G., and Grafton, S. T. (2015). Out of control: diminished prefrontal activity coincides with impaired motor performance due to choking under pressure. *Neuroimage* 105, 145–155. doi: 10.1016/j.neuroimage.2014.10.058

Lin, Y., Mutz, J., Clough, P. J., and Papageorgiou, K. A. (2017). Mental toughness and individual differences in learning, educational and work performance, psychological well-being, and personality: a systematic review. *Front. Psychol.* 8:1345. doi: 10.3389/fpsyg.2017.01345

Liu, R., Walker, E., Friedman, L., Arrington, C. M., and Solovey, E. T. (2020). fNIRS-based classification of mind-wandering with personalized window selection for multimodal learning interfaces. *J. Multimod. User Interfaces*. doi: 10.1007/s12193-020-00325-z

Martens, R. (1977). *Sport Competition Anxiety Test*. Champaign, IL: Human Kinetics Publishers.

Martens, R., Vealey, R. S., and Burton, D. (1990). *Competition Anxiety in Sport*. Champaign, IL: Human Kinetics Publishers.

Memmert, D., Hüttermann, S., Hagemann, N., Loffing, F., and Strauss, B. (2013). Dueling in the penalty box: evidence-based recommendations on how shooters and goalkeepers can win penalty shootouts in soccer. *Int. Rev. Sport Exerc. Psychol.* 6, 209–229. doi: 10.1080/1750984X.2013.811533

Meyer, T., Smeets, T., Giesbrecht, T., Quaedflieg, C. W., Smulders, F. T., Meijer, E. H., et al. (2015). The role of frontal EEG asymmetry in post-traumatic stress disorder. *Biol. Psychol.* 108, 62–77. doi: 10.1016/j.biopsycho.2015.03.018

Molavi, B., and Dumont, G. A. (2012). Wavelet-based motion artifact removal for functional near-infrared spectroscopy. *Physiol. Meas.* 33, 259–270. doi: 10.1088/0967-3334/33/2/259

Molavi, B., May, L. A., Gervain, J., Carreiras, M., Werker, J. F., and Dumont, G. A. (2014). Analyzing the resting state functional connectivity in the human language system using near infrared spectroscopy. *Front. Hum. Neurosci.* 7:921. doi: 10.3389/fnhum.2013.00921

Nguyen, T., Babawale, O., Kim, T., Jo, H. J., Liu, H., and Kim, J. G. (2018). Exploring brain functional connectivity in rest and sleep states: a fNIRS study. *Sci. Rep.* 8:16144. doi: 10.1038/s41598-018-33439-2

Nosrati, R., Vesely, K., Schweizer, T. A., and Toronov, V. (2016). Event-related changes of the prefrontal cortex oxygen delivery and metabolism during driving measured by hyperspectral fNIRS. *Biomed. Opt. Express* 7, 1323–1335. doi: 10.1364/BOE.7.001323

Roberts, L. J., Jackson, M. S., and Grundy, I. H. (2017). Choking under pressure: illuminating the role of distraction and self-focus. *Int. Rev. Sport Exerc. Psychol.* 12, 49–69. doi: 10.1080/1750984X.2017.1374432

Sarason, I. G. (1988). Anxiety, self-preoccupation and attention. *Anxiety Res.* 1, 3–7. doi: 10.1080/10615808808248215

Scholkmann, F., Spichtig, S., Muehlemann, T., and Wolf, M. (2010). How to detect and reduce movement artifacts in near-infrared imaging using moving standard deviation and spline interpolation. *Physiol. Meas.* 31, 649–662. doi: 10.1088/0967-3334/31/5/004

Schweizer, T., Kan, K., Hung, Y., Tam, F., Naglie, G., and Graham, S. (2013). Brain activity during driving with distraction: an immersive fMRI study. *Front. Hum. Neurosci.* 7:53. doi: 10.3389/fnhum.2013.00053

Silveira, R., Prado, R. C. R., Brietzke, C., Coelho-Júnior, H. J., Santos, T. M., Pires, F. O., et al. (2019). Prefrontal cortex asymmetry and psychological responses to exercise: a systematic review. *Physiol. Behav.* 208:112580. doi: 10.1016/j.physbeh.2019.112580

Singh, A. K., and Dan, I. (2006). Exploring the false discovery rate in multichannel NIRS. *Neuroimage* 33, 542–549. doi: 10.1016/j.neuroimage.2006.06.047

Smith, R., Smoll, F., Cumming, S., and Grossbard, J. (2006). Measurement of multidimensional sport performance anxiety in children and adults: the sport anxiety scale-2. *J. Sport Exerc. Psychol.* 28, 479–501. doi: 10.1123/jsep.28.4.479

Smith, R. E., Smoll, F. L., and Schutz, R. W. (1990). Measurement and correlates of sport-specific cognitive and somatic trait anxiety: the sport anxiety scale. *Anxiety Res.* 2, 263–280. doi: 10.1080/08917779008248733

Taelman, J., Vandeput, S., Spaepen, A., and Van Huffel, S. (2009). "Influence of mental stress on heart rate and heart rate variability," in *4th European Conference of the International Federation for Medical and Biological Engineering*, eds J. Vander Sloten, P. Verdonck, M. Nyssen, and J. Haueisen (Berlin; Heidelberg: Springer Berlin Heidelberg), 1366–1369. doi: 10.1007/978-3-540-89208-3_324

Udina, C., Avtzi, S., Durduran, T., Holtzer, R., Rosso, A. L., Castellano-Tejedor, C., et al. (2020). Functional near-infrared spectroscopy to study cerebral hemodynamics in older adults during cognitive and motor tasks: a review. *Front. Aging Neurosci.* 11:367. doi: 10.3389/fnagi.2019.00367

Veale, J. F. (2013). Edinburgh handedness inventory–short form: a revised version based on confirmatory factor analysis. *Laterality* 19, 164–177. doi: 10.1080/1357650X.2013.783045

Wilson, M. R., Wood, G., and Vine, S. J. (2009). Anxiety, attentional control, and performance impairment in penalty kicks. *J. Sport Exerc. Psychol.* 31, 761–75. doi: 10.1123/jsep.31.6.761

Wolf, S., Brölz, E., Keune, P. M., Wesa, B., Hautzinger, M., Birbaumer, N., et al. (2015). Motor skill failure or flow-experience? Functional brain asymmetry and brain connectivity in elite and amateur table tennis players. *Biol. Psychol.* 105, 95–105. doi: 10.1016/j.biopsycho.2015.01.007

Wolf, S., Brölz, E., Scholz, D., Ramos-Murguialday, A., Keune, P., Hautzinger, M., et al. (2014). Winning the game: brain processes in expert, young elite and amateur table tennis players. *Front. Behav. Neurosci.* 8:370. doi: 10.3389/fnbeh.2014.00370

Wood, G., and Wilson, M. R. (2010). A moving goalkeeper distracts penalty takers and impairs shooting accuracy. *J. Sports Sci.* 28, 937–946. doi: 10.1080/02640414.2010.495995

Wood, R. (2017a). *About the Sport Competition Anxiety Test (SCAT)*. Available online at: https://www.topendsports.com/psychology/scat.htm

Wood, R. (2017b). *The Sport Anxiety Scale (SAS)*. Available online at: https://www.topendsports.com/psychology/sas.htm

Yoon, J. H., Curtis, C. E., and D'Esposito, M. (2006). Differential effects of distraction during working memory on delay-period activity in the prefrontal cortex and the visual association cortex. *Neuroimage* 29, 1117–1126. doi: 10.1016/j.neuroimage.2005.08.024

Yu, R. (2015). Choking under pressure: the neuropsychological mechanisms of incentive-induced performance decrements. *Front. Behav. Neurosci.* 9:19. doi: 10.3389/fnbeh.2015.00019

Yücel, M. A., Selb, J., Cooper, R. J., and Boas, D. A. (2014). Targeted principle component analysis: a new motion artifact correction approach for near-infrared spectroscopy. *J. Innov. Opt. Health Sci.* 7:1350066. doi: 10.1142/S1793545813500661

Zhu, F., Poolton, J., Wilson, M., Maxwell, J., and Masters, R. (2011). Neural co-activation as a yardstick of implicit motor learning and the propensity for conscious control of movement. *Biol. Psychol.* 87, 66–73. doi: 10.1016/j.biopsycho.2011.02.004

Zhu, F. F., Maxwell, J. P., Hu, Y., Zhang, Z. G., Lam, G. W., Poolton, J. M., et al. (2010). EEG activity during the verbal-cognitive stage of motor skill acquisition. *Biol. Psychol.* 84, 221–227. doi: 10.1016/j.biopsycho.2010.01.015

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

frontiers in Computer Science

ORIGINAL RESEARCH
published: 14 June 2021
doi: 10.3389/fcomp.2021.661224

Evaluating a Novel P300-Based Real-Time Image Ranking BCI

Ngadhnjim Sutaj[1,2], Martin Walchshofer[1], Leonhard Schreiner[1,3], Luca Turchet[2], Harald Pretl[3] and Christoph Guger[1]*

[1]g.tec Guger Technologies OG/g.tec Medical Engineering GmbH, Schiedlberg, Austria, [2]Department of Information Engineering and Computer Science, University of Trento, Trento, Italy, [3]Institute for Integrated Circuits, Johannes Kepler University, Linz, Austria

Brain–computer interfaces (BCIs) establish communication between a human brain and a computer or external devices by translating the electroencephalography (EEG) signal into computer commands. After stimulating a sensory organ, a positive deflection of the EEG signal between 250 and 700 ms can be measured. This signal component of the event-related potential (ERP) is called "P300." Numerous studies have provided evidence that the P300 amplitude and latency are linked to sensory perception, engagement, and cognition. Combining the advances in technology, classification methods, and signal processing, we developed a novel image ranking system called the Unicorn Blondy Check. In this study, the application was tested on 21 subjects using three different visual oddball paradigms. Two consisted of female faces and gray-scale images, while the third test paradigm consisted of familiar and unfamiliar faces. The images were displayed for a duration of 150 ms in a randomized order. The system was trained using 50 trials and tested with 30 trials. The EEG data were acquired using the Unicorn Hybrid Black eight-channel BCI system. These synchronized recordings were analyzed, and the achieved classification accuracies were calculated. The EEG signal was averaged over all participants and for every paradigm separately. Analysis of the EEG data revealed a significant shift in the P300 latency dependent on the paradigm and decreased amplitude for a lower target to non-target ratio. The image ranking application achieved a mean accuracy of 100 and 95.5% for ranking female faces above gray-scale images with ratios of 1:11 and 5:11, respectively. In the case of four familiar faces to 24 unfamiliar faces, 86.4% was reached. The obtained results illustrate this novel system's functionality due to accuracies above chance levels for all subjects.

Keywords: BCI, EEG, P300, neuromarketing, image-ranking, VEP, cognition, familiarity

OPEN ACCESS

Edited by:
Anton Nijholt, University of Twente, Netherlands

Reviewed by:
Tomasz Maciej Rutkowski, RIKEN Center for Advanced Intelligence Project (AIP), Japan
Felix Putze, University of Bremen, Germany

****Correspondence:***
Ngadhnjim Sutaj
sutaj@gtec.at

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Computer Science

Received: *30 January 2021*
Accepted: *12 May 2021*
Published: *14 June 2021*

Citation:
Sutaj N, Walchshofer M, Schreiner L, Turchet L, Pretl H and Guger C (2021) Evaluating a Novel P300-Based Real-Time Image Ranking BCI. Front. Comput. Sci. 3:661224. doi: 10.3389/fcomp.2021.661224

1 INTRODUCTION

Brain–computer interfaces (BCIs) establish communication between a human brain and a computer or external devices. The BCI translates information of electrophysiological signals measured from the scalp *via* electroencephalography (EEG) or directly from the cortex using electrocorticography into computer commands (Wolpaw et al., 2002). EEG-based BCIs provide an inexpensive, straightforward, and noninvasive method for studying neural activities. Therefore, they are widely used in research environments and commercial applications. Principles on which BCIs rely are motor-imagery, slow waves, steady-state visual evoked potentials (VEPs), and evoked potentials (Schomer and Silva, 2010; Nicolas-Alonso and Gomez-Gil, 2012). Pfurtscheller (2001) was one of the first to show a correlation between the EEG signal and imagining body movement called event-related synchronization and desynchronization.

In recent years, there has been growing interest in studying cognitive neuroscience with the help of EEG-based BCIs. A robust methodology to study cognitive processing is event-related potentials (ERPs) (Woodman, 2010). When exposed to specific events or stimuli, brain structures generate these ERPs. The type of evoked potential depends on the presented stimulus. Whether auditory, olfactory, or visual evoked potentials, they all result in the corresponding ERP (Sato et al., 1996). VEPs, a subcategory of ERPs, find widespread use in BCIs, where they enable information exchange without physical inputs (Nicolas-Alonso and Gomez-Gil, 2012; Zhang et al., 2012). While visual stimuli are mainly used, auditory and tactile stimuli have also been useful in BCI applications (Gao et al., 2014; Lugo et al., 2014).

Specific components within the ERP, called cognitive negative variation and P300, were first discovered by Walter et al. (1964). P300, in this context, represents the positive deflection of the ERPs amplitude about 250–700 ms after a novel stimulus compared to the previous stimuli is presented (Donchin and Smith, 1970; Rugg and Coles, 1995). Numerous studies have provided evidence that P300 is linked to sensory perception, engagement, and cognition (Woodman, 2010; Chen et al., 2020). These discoveries have made P300 an essential parameter in neuroscience. Aiming to leverage the P300 component, the standard paradigm used in P300-based BCIs is the “oddball” paradigm (Luck, 2014). The paradigm relies on the fact that an unexpected stimulus triggers a higher P300 amplitude than a reference stimulus. On the other hand, results from several studies have suggested that unexpected or rare stimuli always produce a P300 (Groppe et al., 2011), (Chen et al., 2020). Also, it has been observed that a lower probability of a target stimulus results in a higher P300 amplitude (Polich et al., 1996), (Duncan-Johnson and Donchin, 1977).

Farwell and Donchin (1988) proposed the first P300-based BCI. A six-by-six matrix composed of letters representing different commands was presented to the subjects. Then the columns and rows of the matrix were flashed rapidly in a random order, and the subjects were asked to focus on the desired letter and count the number of flashes. Only flashes of rows and columns containing the desired letter evoke P300 potentials. Commonly used in BCI research for this kind of application is the term P300-speller. Besides the row/column speller mentioned above, single-character spellers have also been investigated. The basic idea with single-character spellers is that a lower probability of the target occurring triggers a higher P300 response. However, no higher P300 amplitude or more reliable control of single-character spellers compared to spellers could be confirmed in prior studies (Guger et al., 2009).

Several feature extraction and classification procedures have been investigated to enhance the performance of P300-based BCIs (Krusienski et al., 2008). Methods such as linear discriminant analysis (Guger et al., 2009), stepwise linear discriminant analysis (Sellers and Donchin, 2006), support vector machines (Thulasidas et al., 2006), and matched filtering (Serby et al., 2005) have been utilized in P300-based BCIs. The field is maturing with the adoption of new machine learning algorithms that reduce the amount of training data required to achieve sufficient classification accuracy, such as the time-variant Linear discriminant analysis proposed by Gruenwald et al. (2019). Complementing these advances, researchers have also focused on improving P300-based BCIs' slow communication rates (Martens et al., 2009; Takano et al., 2009; Mugler et al., 2010).

Relevant to the application of P300-based BCIs in the nonclinical area are studies that hint at the P300 latency and shape change due to the cognitive processes associated with the stimulus processing. Specifically, it is suggested that the latency is related to the time it takes the person to process the oddball stimulus and that the P300 component of the ERP might be linked to sensory perception, engagement, and cognition (McCarthy and Donchin, 1981; Magliero et al., 1984; Rugg and Coles, 1996; Comerchero and Polich, 1999; Polich, 2007). Although a large body of research exists, suggesting the P300 shape, amplitude, and latency correlate with the underlying stimuli, only a few applications have adopted this groundwork for commercial and clinical use. Mainly spelling systems such as the intendiX system (Guger et al., 2016) or the mindBEAGLE BCI for patients with disorders of consciousness (Guger et al., 2017) are currently available.

Leveraging the previously mentioned research and advances, g.tec neurotechnology GmbH has developed the Unicorn Blondy Check. This application is presented in this study and aims to advance neuromarketing and enable new VEP research findings. Based on synchronizing the EEG samples with the image presentation, the system uses the P300 response to assign a score value to the displayed images and ranks the images according to that score value.

This study aims to present and validate a working EEG-based image ranking software. For this purpose, we investigated whether features extracted from a live EEG recording can be used to select and rank images with an accuracy above chance level. Subsequently, the application can be applied as a ranking system to reduce the time for sorting/labeling/selecting numerous images and for further research on the connection between the EEG and cognition.

2 MATERIALS AND METHODS

2.1 Apparatus

The EEG signal was recorded using the Unicorn Hybrid Black system (g.tec neurotechnology GmbH, 2020; Accessed: 2021-27-01). The biosignal amplifier was connected to a personal computer using the integrated Bluetooth interface. The apparatus provides a 24-Bit conversion with a sampling rate of 250 Hz. Eight channels are recorded on the following positions: $\{Fz, C3, Cz, C4, Pz, PO7, Oz, PO8\} = \{CH1, CH2, CH3, CH4, CH5, CH6, CH7, CH8\}$. The channel positions are visualized in **Figure 1**. Ground and reference are

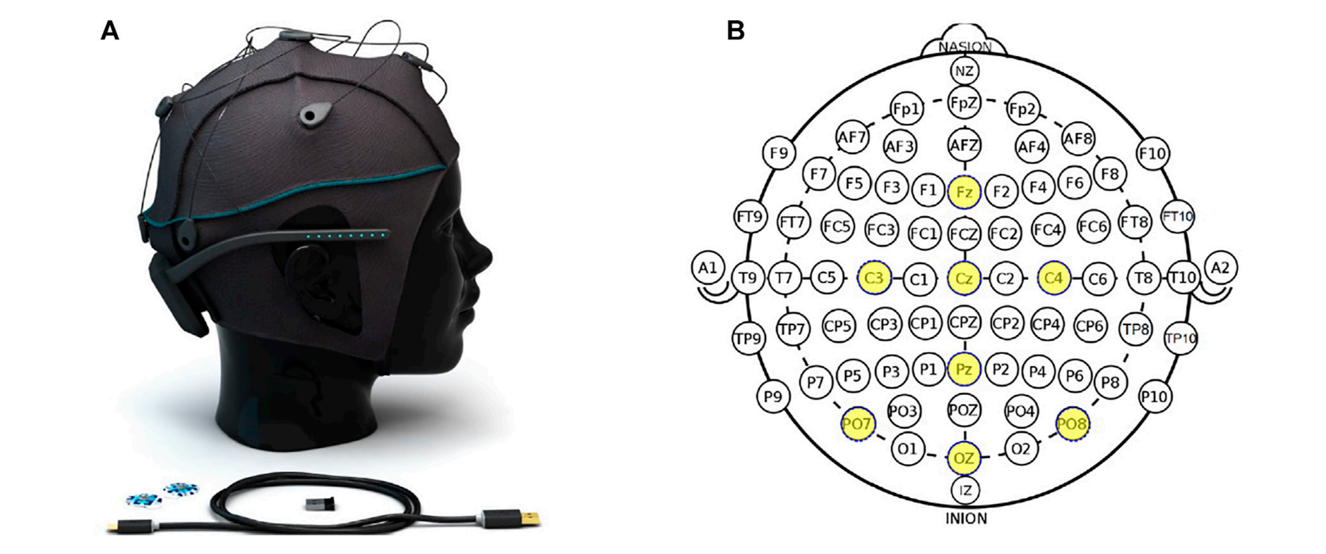


FIGURE 1 | (A) Unicorn Hybrid Black system, **(B)** the electrode positions of the system. Ground and reference are fixed on the mastoids using disposable sticker-based surface electrodes and not visible in the graphic.

TABLE 1 | Composition of the four different paradigms used for the image ranking experiment.

Paradigm name	Target count	Non-target count	Target image	Non-target image
P1	1	11	Female face	Gray-scale
P2	5	11	Female face	Gray-scale
P3	4	24	Familiar face	Unfamiliar face
P4	1	11	Familiar face	Unfamiliar face

placed on the mastoids of the subject using the disposable sticker-based surface electrodes. The data are recorded and stored using the Unicorn Blondy Check application (g.tec neurotechnology GmbH, 2021; Accessed: 2021-27-01).

2.2 Paradigm Design

For the image ranking experiment, different random patterns of images need to be generated to train and test the image ranking system. These patterns are referred to as paradigms P1, P2, P3, and P4. They are essentially oddball paradigms. To create these paradigms, the application's paradigm editor is used. The oddball in P1 and P2 is a colored picture of a female face compared to gray-scale images, while the oddball in P3 and P4 is a familiar face compared to unfamiliar faces. P1 has one target and 11 non-targets, P2 has four targets and 11 non-targets, P3 has four targets and 24 non-targets, and P4 has one target and 11 non-targets. The four different paradigms are summarized in **Table 1**, while the images used for each paradigm are depicted in **Figure 2**. The pictures of the familiar faces are unique for each subject. Thus they are replaced by placeholders-icons in **Figures 2C,D**. The ratio of 1:11 for target to non-target stimuli for the system's calibration was chosen in accordance with P300-based speller systems (Guger et al., 2009). To test the system, paradigms with 1:11 as well as lower ratios of 4:11 and 4:24 were generated. These lower ratios were chosen to investigate if the decreases in P300-amplitude due to lower ratios as suggested by Duncan-Johnson and Donchin (1977) could lead to a decrease in Ranking performance. Regarding the choice of images, faces were chosen because recent work has shown that faces can improve P300 BCI performance (Guger et al., 2016).

2.3 Participants and Procedure

Twenty-one subjects (9 males, 12 females), aged between 22 and 78 (mean age = 35, standard deviation = 14), took part in the study. All participants provided informed consent and were recruited through word-of-mouth. The study was approved by the responsible ethical committee (Ethikkommission des Landes Oberösterreich; Number D-35-16). The tests were conducted in multiple locations, mainly the home of the participants or g.tec company grounds. Special care was taken that the environment was as quiet as possible to keep the distraction to a minimum.

Participants were asked to sit in front of the computer screen. The system was placed on the subject's head. Then, the EEG-cap was connected to the application *via* Bluetooth. The signal-quality was checked using the built-in signal quality feature. The contact impedance (electrode-scalp) was lowered by injecting electrode gel to improve the signal quality. The gel was applied to each of the eight electrodes of the EEG-cap. After this step, the signal quality evaluated by the system was within limits set by the system. The steps described above were done for

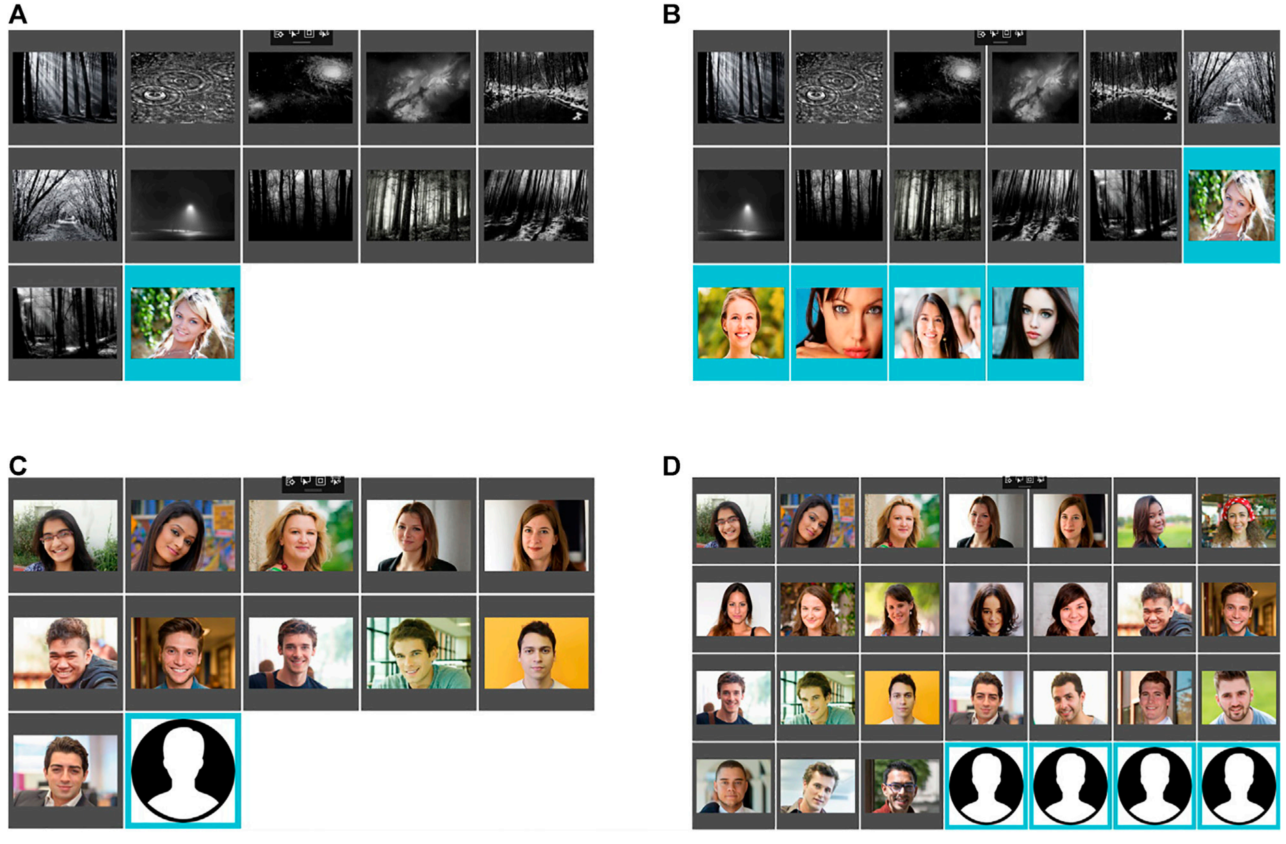


FIGURE 2 | Image sets used for the oddball paradigms: **(A)** Paradigm 1 "P1" containing one female face (target class) and 11 gray-scale images (non-target class), **(B)** Paradigm 2 "P2" containing five female faces (target class) and 11 gray-scale images (non-target class), **(C)** Paradigm 4 "P4" containing one familiar face (target class) and 11 unfamiliar faces (non-target class), **(D)** Paradigm 3 "P3" containing four familiar faces (target class) and 24 unfamiliar faces (non-target class). The pictures of the familiar faces are different for each subject. Thus they are replaced by placeholders-icons in **(C)** and **(D)**.

TABLE 2 | Measurement structure. In part one, the application was trained on the VEP produced by a female face and gray-scale images (P1). In part two, familiar faces and unfamiliar faces (P4) were used for training. P1, P2, and P3 were used for testing. Each image was displayed 150 ms, with the next immediately following the previous. The training paradigm was presented 50 times, and each test paradigm 30 times.

Mode	Paradigm	Trials	Presentation Time (ms)	Number of Images	Duration [s]
Measurement part one					
Training	P1	50	150	12	90
Test 1	P1	30	150	12	54
Test 2	P2	30	150	16	72
Test 3	P3	30	150	28	126
Measurement part two					
Training	P4	50	150	12	90
Test 1	P1	30	150	12	54
Test 2	P2	30	150	16	72
Test 3	P3	30	150	28	126

every subject. Finally, the different test paradigms (**Figure 2**) were loaded. The EEG was recorded throughout the entire measurement. After completing the training, the test paradigms were loaded. The measurement procedure can be segmented into two parts, with each having training and a test phase, as described below.

2.3.1 Part One

During the first part of the measurement, the system was trained using P1 (one female face and 11 gray-scale images) as a training paradigm (**Figure 2**). The training consisted of 50 trials resulting in 50 target stimuli and 550 non-target stimuli, which amounted to a training time of 90 s. After completing the training, the system was tested using the paradigms P1, P2, and P3. Each test consisted of 10 trials and was performed three times resulting in 30 test trials for each paradigm. The measuring sequence of part one can be seen in **Table 2**.

2.3.2 Part Two

During the second part of the measurement, P4 (one familiar face, 11 unfamiliar faces) was loaded for the training paradigm (see **Figure 2**). Analogous to part one, the training consisted of 50 trials resulting in 50 target and 550 non-target stimuli, which amounted to a training time of 90 s. The trained system was tested using the paradigms P1, P2, and P3 analogously to part one. Each test consisted of 10 trials and

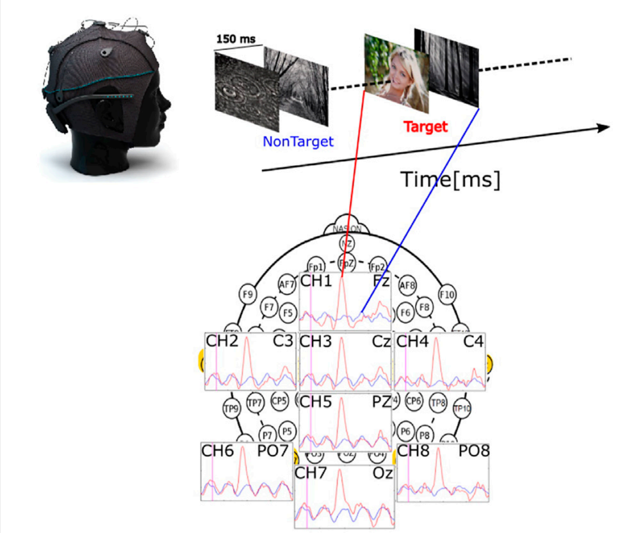


FIGURE 3 | Schematic overview of the stimuli presentation with an example response averaged over the training trials of one subject with regard to electrode placement. The evoked potential is drawn in red for the target stimulus and blue for the non-target stimulus.

was performed three times resulting in 30 test trials for each paradigm. The measuring sequence of part two can be seen in **Table 2**.

2.3.3 Stimulus Presentation Parameters

Each image was shown for a duration of 150 ms immediately followed by the next in the trial. No dark screen was displayed between the pictures, as shown in **Figure 3**. One trial consisted of 12 pictures for P1 and P4, 16 for P2, and 28 for P3, as summarized in **Table 1**. The application randomizes the order of the images during the trial. The training phase consisted of 50 and the testing phase consisted of 30 trials.

2.4 Data Processing and Classification

The EEG signals are acquired sample-wise from the Unicorn Hybrid Black system *via* the Bluetooth connection. The application then synchronizes the EEG samples with the image presentation. To reduce the mains line interference (50 Hz EU/60 Hz United States) the EEG samples are digitally filtered using a second-order 50 Hz Butterworth Notch filter followed by a second-order 60 Hz Butterworth Notch filter. Finally, a second-order Butterworth bandpass filter with a band ranging from 0.5 to 30 Hz is applied to improve the signal-to-noise ratio. After the single samples are filtered, 1,500 ms-epochs are created using 100 ms before and 1,400 ms after the stimulus onset. For feature extraction, each epoch is baseline corrected using the 100 ms before the stimulus. The epoch is downsampled by a factor of 1:12 and moving average filtered using a window size of 3. The mentioned procedure results in a feature with a length of 248 (31 samples * 8 channels). Eye blinks are not corrected/removed and there was no artifact rejection algorithm used.

2.5 Ranking

The target and non-target features recorded during the training phase are used to fit a time-variant linear discriminant analysis model. Cai et al. (2013) have already proven that a linear regression in the LDA subspace is mathematically equivalent to a low-rank linear regression. Based on this knowledge, the features generated by subsequent images during the ranking are projected into the LDA subspace and the resulting distance (score value) is used to rank the new images.

2.6 Data Evaluation

2.6.1 Electroencephalography Study

The system provides the user with the possibility of recording the stimulus synchronized EEG data during the experiment. The EEG recordings were averaged over the 30 trials and the 21 participants to visualize the recorded visual evoked potentials measured during the experiment. This averaging was done for every paradigm using the MATLAB R2020a software. Additionally, the P300 latency and amplitude were marked by means of maxima detection on the averaged waveforms for each paradigm. This results in one P300 amplitude and latency per electrode and paradigm. The amplitude and latency differences are examined for significance using a paired sample t-test.

2.6.2 Ranking Accuracy

The system produces score values for every displayed image. These values were compared to evaluate the image ranking performance. If the target class image had a higher score value than the non-target class's images, the result was determined to be a successful ranking. The values for every image were averaged and compared again for each trial. This evaluation was done for 1 – 30 trials and further averaged over every subject (n = 21). Based on this information, a ranking accuracy could be estimated. This accuracy corresponds to the true positive rate described as the ratio between correct classifications and the total number of classifications.

3 RESULTS

3.1 Electroencephalography Study

Figures 4, **5** depict the calculated difference between target and non-target stimuli averaged over 21 participants. **Figure 4A** illustrates the visual evoked potential produced by paradigm P1. **Figure 4B** shows the visual evoked potential produced by paradigm P2. **Figure 5A** corresponds to the Visual evoked potential produced by paradigm P3. **Figure 5B** illustrates the visual evoked potential produced by paradigm P4. The latency (ms) and the amplitude (μV) of the P300 are marked for each electrode position within the individual graphs. **Table 3** lists the latency and amplitude derived from **Figures 4**, **5** for all four paradigms and all eight electrode positions.

Paradigms with female faces and gray-scale images produce the following mean P300 latencies: For P1, the latency is 233 ms (Fz-Pz) and 222 ms (PO7-PO8). For P2, it is 232 ms (Fz-Pz) and 226 ms (PO7-PO8). Changing the paradigm to

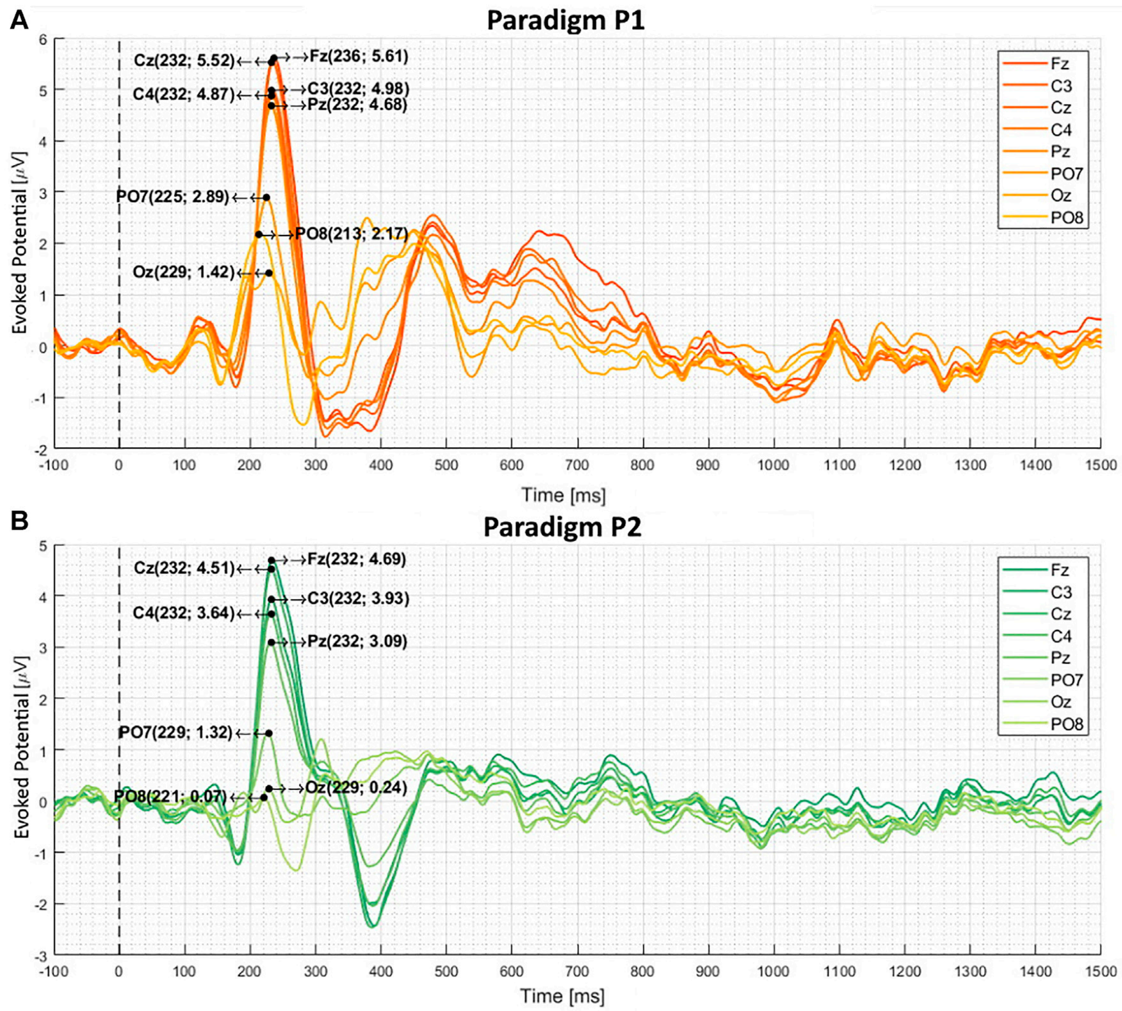


FIGURE 4 | The calculated difference between target and non-target stimuli averaged over 21 subjects **(A)** and **(B)** depicts the visual evoked potential produced by paradigms P1 and P2, respectively. These two paradigms had female faces as targets and gray-scale images as non-targets. The black dotted line marks the stimulus presentation. The P300 latency and amplitude are noted in the brackets [(ms), (μV)].

familiar and unfamiliar faces, the latency increases to 340 ms (Fz-Pz) and 356 ms (PO7-PO8) for P4 and 314 ms (Fz-Pz) and 307 ms (PO7-PO8) for P3. These values can be seen in **Table 3**.

The P300 amplitude evoked by paradigms with different target to non-target ratios decreases from 5.13μV (Fz-Pz) and 2.16μV (PO7-PO8) for P1 (ratio = 1:11) to 3.97μV (Fz-Pz) and 0.54μV (PO7-PO8) for P2 (ratio = 5:11). Similarly, the amplitude decreases from 3.70μV (Fz-Pz) and 2.65μV (PO7-PO8) for P4 (ratio = 1:11) to 2.86μV (Fz-Pz) and 1.86μV (PO7-PO8) for P3 (ratio = 4:24). These values can be seen in **Table 3**.

3.2 Image Ranking Performance

In this section, the accuracy results for the first and second parts of the experiment are presented. For part one, the application was trained on the EEG features produced by one female face (target) and 11 gray-scale images (non-targets). For part two, the system was trained on the EEG features produced by one familiar face (target) and 11 unfamiliar faces (non-targets). The training consisted of 50 trials. Paradigms P1, P2, and P3 were used for testing with 30 test trials each. **Figure 6A** depicts the accuracy for part one averaged over all subjects. Analogously **Figure 6B** shows the accuracy for part two averaged over all participants. **Figures 6C,D** depict the median accuracy as well as the range between the 25th and 75th percentiles (shaded area) for parts one and two respectively over all participants.

The results shown in **Figures 6A,B** are summarized in **Table 4** by additional averaging over all trials. The achieved accuracy after 30 trials is listed in the column marked with "Final" and represents the ranking accuracy at 30 trials. For part one of the measurements, the application reached an online accuracy of 100, 95.5, and 14.8% for P1, P2, and P3, respectively. When switching the training paradigms the achieved accuracy changes to 38.6, 34.6, and 86.4% for P1, P2, and P3 in part two of the measurements.

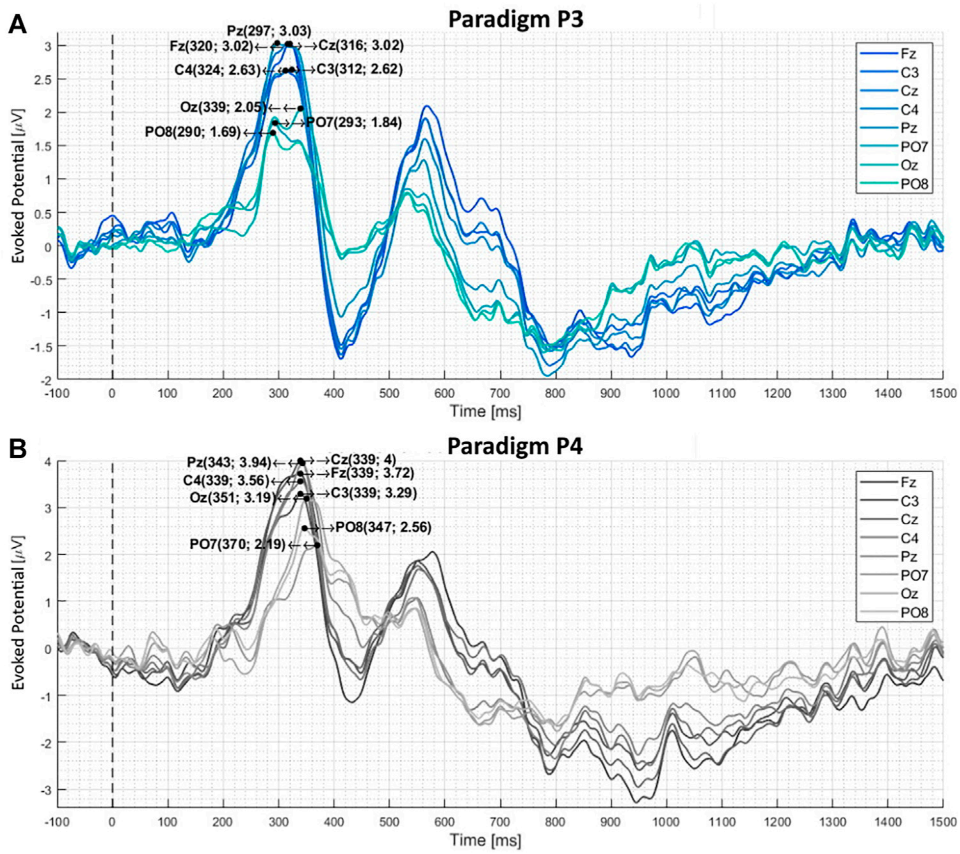


FIGURE 5 | The calculated difference between target and non-target stimuli averaged over 21 subjects **(A)** and **(B)** depicts the visual evoked potential produced by paradigms P4 and P3, respectively. These two paradigms had familiar faces as targets and unfamiliar faces as non-targets. The black dotted line marks the stimulus presentation. The P300 latency and amplitude are noted in the brackets [(ms), (μV)].

TABLE 3 | P300 latency (*ms*) and amplitude (*μV*) taken from **Figures 4**,**5** is listed for every paradigm (P1, P2, P3, and P4) and all electrode positions averaged for all subjects. The average of Fz, C3, Cz, C4, Pz, and PO7, Oz, PO8 is calculated separately and listed.

Electrode	Female face/Gray-scale		Familiar/unfamiliar faces	
Position	**P1**	**P2**	**P4**	**P3**
Fz	236 (5.61)	232 (4.69)	339 (3.72)	320 (3.02)
C3	232 (4.98)	232 (3.93)	339 (3.29)	312 (2.62)
Cz	232 (5.52)	232 (4.51)	339 (4.00)	316 (3.02)
C4	232 (4.87)	232 (3.64)	339 (3.56)	324 (2.63)
Pz	232 (4.68)	232 (3.09)	343 (3.94)	297 (3.03)
PO7	225 (2.89)	229 (1.32)	370 (2.19)	293 (1.84)
Oz	229 (1.42)	229 (0.24)	351 (3.19)	339 (2.05)
PO8	213 (2.17)	221 (0.07)	347 (2.56)	290 (1.69)
[Fz,C3,Cz,C4,Pz]Avg	233 (5.13)	232 (3.97)	340 (3.70)	314 (2.86)
[PO7,Oz,PO8]Avg	222 (2.16)	226 (0.54)	356 (2.65)	307 (1.86)

4 DISCUSSION

The main objective of this study was to evaluate and present a novel visual evoked potential-based image ranking BCI. In the first step, the recorded EEG signals are discussed. Second, the achieved accuracy will be addressed.

4.1 Electroencephalography Study

The P300 peak (averaged over all participants and trials) is visible in **Figures 4**, **5** for all paradigms (P1, P2, P3, P4) at the electrode positions $\{Fz, C3, Cz, C4, Pz, PO7, Oz, PO8\}$.

4.1.1 P300 and Task Complexity

When comparing the average target P300 response time of female/gray-scale (P1, P2) to familiar/unfamiliar faces (P3, P4) paradigms, a significant ($p < 0.001$) increase in the mean P300 latency can be seen at all electrode positions, visible in **Figures 4**, **5** and listed in **Table 3**. The delay could be caused by the task complexity associated with difficult stimulus discrimination, as Polich (2007) suggested. This complexity might be the case when distinguishing familiar from unfamiliar faces. In contrast, distinguishing a colored female face image from a gray-scale photo provides less of a challenge. Examining the average amplitude differences at the electrode positions Fz, C3, Cz, C4, and Pz analogously to the response times, a decrease for P4-P1 and P3-P2 is visible in **Figures 4**, **5**. Only the comparison between P4-P1 should be considered, as both paradigms have an equal target

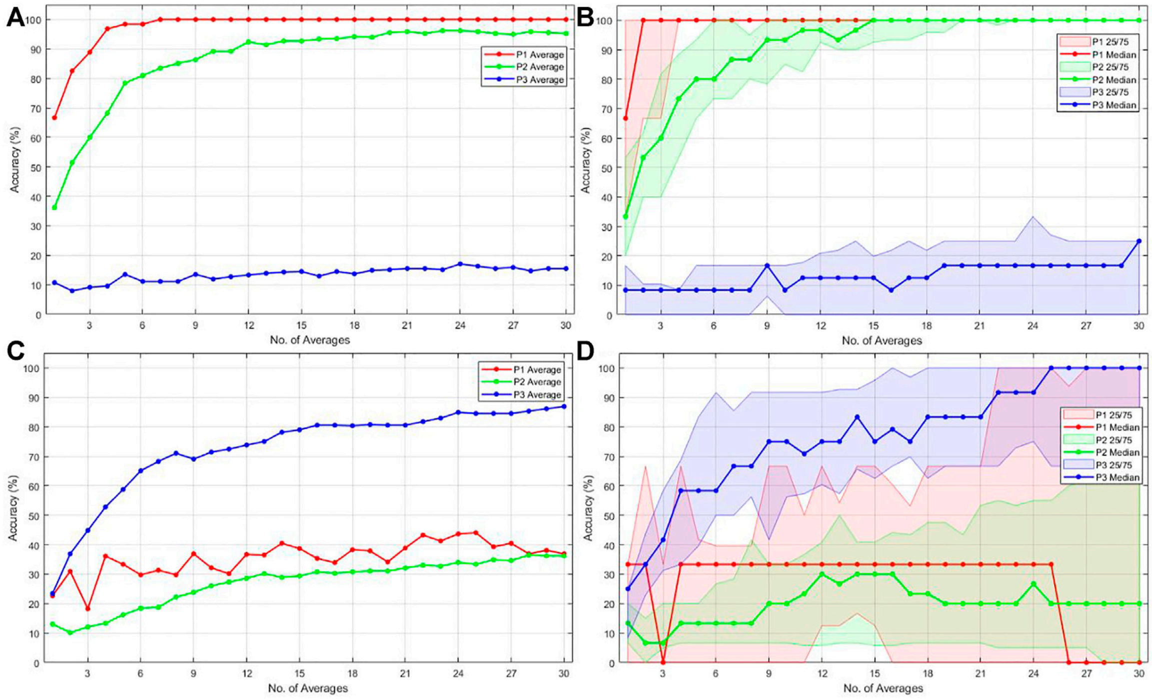


FIGURE 6 | Accuracy results for 1 to 30 test trials and paradigms P1, P2, and P3 for 21 subjects. Measurement setup part one **(A)**, **(C)** (Training paradigm P1) and part two **(B)**, **(D)** (Training paradigm P4). On the left side **(A)** and **(B)** depict the average accuracy. On the right side **(C)** and **(D)** depict the median accuracy with the shaded area representing the 25th (Q1) and 75th (Q3) percentiles.

TABLE 4 | Ranking performance accuracy results, taken from **Figure 6 (A)** for part one and **Figure 6 (B)** for part two. The mean is an average of 1–30 test trials. The median is calculated of trials 1 to 30, and the final accuracy value is the achieved accuracy after 30 trials. Additionally, the accuracy difference between the two measurement parts is listed and is calculated by subtracting the respective accuracy value of part one (training with P1) from part two (training with P4).

Test	ACC [%] Figure 6A			ACC [%] Figure 6B			ACC [%] difference		
	Final	Mean	Median	Final	Mean	Median	Final	Mean	Median
P1	100	97.8	100	38.6	36.4	37.3	– 61.4	– 61.4	– 62.7
P2	95.5	86.9	93.3	34.6	37.3	28.9	– 60.9	– 49.6	– 64.6
P3	14.8	12.9	13.5	86.4	72.0	78.7	+ 71.6	+59.1	+65.2

to non-target ratios of (1:11), while P2 and P3 have a ratio of (5:11) and (4:24), respectively. This argument is based on the fact that the P300 amplitude induced by an oddball paradigm depends on the oddball stimulus's rarity as described by (Luck, 2014), (Wolpaw, 2012), and (Duncan-Johnson and Donchin, 1977). The slightly higher amplitude of P1, compared to P4, could be caused by the stronger contrast between the target and non-target images. Consequently, the more pronounced difference could magnify the resulting P300 response, even though they have the same target to non-target ratio.

4.1.2 P300 and Target to Non-Target Ratio

As early as 1977, Duncan-Johnson and Donchin (1977) observed that the P300 amplitude correlates with the associated oddball a priori probability. Simply put, the less likely the oddball (target) stimulus is to occur, the higher the P300 amplitude. This observation is also visible in the measurement data. A significant ($p < 0.001$) decrease in amplitude is visible for the ratios of 1:11(P1) and 5:11 (P2) listed in **Table 3**. Analogously, a reduction in the P300 amplitude for 1:11(P4) and 4:24 (P3) is also visible.

To sum up **section 4.1.2** and **section 4.1.1**, the P300 amplitude is highly dependent on the oddball frequency for any given paradigm. In contrast, the P300 latency and waveform shape are dependent on the task complexity of the specific paradigm class.

4.2 Classification Accuracy

The mean and median accuracy for the first and second parts of the measurement listed in **Table 4** does not represent the achieved accuracy for 30 consecutive trials but rather the averaged accuracy for 1 to 30 trials. The achieved average accuracy for 30 trials is also listed in **Table 4** in the "Final" accuracy columns. The EEG data were processed and classified online. No cross-validation was performed as the randomization constituted by cross-validation would lead to an over-optimistic accuracy estimation, as Wolpaw (2012) suggested. The combination of these two aspects led to a realistic accuracy assessment, as listed in **Table 4**. The final average accuracy after 30 trials combined with the mean and median can serve the purpose for a relative comparison between the test paradigms and, more importantly, between different training paradigms' effects on the ranking performance. The average and median achieved online accuracy for a specific test trial count for all subjects is visible in **Figure 6**. The average rather than the median accuracy was discussed further, since the average is strongly affected by subjects where the application performed worse. Therefore, the average provides a more practical estimation.

4.3 Measurement for Part One

Figure 6A shows that an average accuracy of 100% for test paradigm P1 was reached after seven trials. Mean accuracy of 100% translates to the target images being ranked on top for all subjects after seven trials. The mean accuracy for test paradigm P2 never reached 100%. However, a satisfactory accuracy of 90% was eventually achieved at 10 test trials and increased to over 95% for 20-30 test trials. In paradigm P2, all five female face pictures must be ranked on top to be considered successful. This condition makes the estimate of accuracy even more conservative. An accuracy of 95.5% means that the application ranked all five female face pictures on top for more than 95.5% of the time for all subjects after 30 trials. The reason why even at 30 test trials, the accuracy for P2 never reached 100% could be that the training was performed on P1, and the target visual evoked potentials produced by P1 and P2 slightly differ in amplitude, as discussed in **section 4.1**. The visual evoked potential produced by P3 not only varies in amplitude but also significantly in latency and waveform, as discussed in **section 4.1**. This difference in waveform shape and amplitude could explain why the mean accuracy for test paradigm P3 never reached a good value with only 14.8% at 30 trials.

4.4 Measurement for Part Two

Using one familiar face and 11 unfamiliar faces (P4) as a training paradigm in the second part of the measurement results in an accuracy change for all test paradigms as expected when considering the EEG study results. The accuracy of test paradigms P1 and P2 decreased by 61.4 and 60.9%, respectively. The achieved accuracy at 30 test trials are depicted in **Figure 6B** and listed in **Table 4**. They are 38.6 and 34.6% for P1 and P2. The mean and median accuracy estimation for 1–30 trials from **Table 4** shows a decrease by 61.4 and 62.7% for the ranking of test paradigm P1. Similarly, the mean and median accuracy estimation for test paradigm P2 decreased by 49.6 and 64.6%, respectively.

Confirming expectations that training the application on a paradigm consisting of a familiar face and unfamiliar faces, the achieved accuracy after 30 trials increases by 71.6–86.4% for paradigm P3. Similarly, the mean and median test accuracy for 1 to 30 trials for paradigm P3 increased by 59.1 and 65.2%. This is visible in Figure 6B and is listed in **Table 4**. To rephrase these results into a ranking performance, this means that all four familiar faces are ranked on top of the 24 unfamiliar faces by the application 86.4% of the time.

For paradigms P3 and P4, the subjects had to provide pictures of familiar faces, but these were not always of equivalent quality/brightness to the stock images. The quality variation could be why the accuracy in part two of the measurement for test paradigm P3 was not as high as the accuracy for P2 in part one. The quality of the images could be a significant factor due to the short presentation time of 150 ms.

In summary, it is plausible to state that the choice in training paradigm and the ratio of targets to non-targets for training and test paradigms are essential for the system's performance. These choices change the resulting classification accuracy significantly. For example, training the application on P1 (female face/gray-scale) will not yield satisfactory ranking performance for test paradigm P3 (familiar/unfamiliar faces). Conversely, training the system on P4 (familiar/unfamiliar faces) will lead to a significantly better ranking performance for similar paradigms P3 (familiar/unfamiliar faces) but decrease the ranking performance for P1 and P2 (female face/gray-scale).

5 CONCLUSION AND FUTURE WORK

This study focused on evaluating and presenting the image ranking software Unicorn Blondy Check. The software is based on the P300 component of the visually evoked potential. The evaluation was done by testing different visual paradigms and training the system to detect such differences. By studying the EEG recordings, it was possible to notice the differences in the visually evoked potential for each paradigm. These observed differences meet the expectations suggested by previous research (Donchin and Smith, 1970; McCarthy and Donchin, 1981; Magliero et al., 1984; Comerchero and Polich, 1999; Polich, 2007). Depending on the paradigm used for training, average accuracy of 100% for P1, 95.5% for P2, and 86.4% for P3 were achieved. Selecting the inappropriate training paradigm resulted in decreased ranking performance as expected from the EEG study findings.

Our results showcase the software's image ranking capabilities. Even complex tasks such as ranking photos of familiar faces higher than unfamiliar faces yielded satisfactory results. Nevertheless, several open questions worth investigating in future research remain, which will improve the system's performance. In future work, we plan to address the following:

- To change the stimulus presentation parameters (presentation-time, inter-stimuli-time, presentation order).
- To try different paradigms to investigate what is possible to classify (e.g., classifying affection, determining advertisements with the highest impact, or finding concealed information using paradigms containing crime-related images).
- To measure the accuracy of dry EEG acquisition.

The application areas of such a system are numerous. For researchers, this system can provide an easy-to-use tool for further investigation of the visual evoked potentials. It provides not only classification capabilities but also logs stimulus-synchronized EEG data. The build-in paradigm editor enables rapid testing. In neuromarketing, this system can supplement the existing methods for determining advertisement's impact. With it, one can directly compare the brain's responses to successful and unsuccessful advertising. The software can be used to check whether the image ranking presented in this study also works with paradigms that do not contain faces, for example, different shoe models and dishes or holiday destinations. However, to do this, one must construct calibration paradigms containing images of objects that the test person prefers in comparison with the others. Then sets of unknown images can be examined for preferences or lack of preferences. Researchers in the area of polygraphy and concealed information testing could also benefit from such an easy-to-use system. Finally, it can be a great hands-on learning tool for the everyday person and students interested in BCI as it is more affordable than other EEG systems on the market.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion 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 Ethikkommission des Landes Oberosterreich; Number D-35-16. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

NS developed the methods, was involved in data acquisition, performed data processing, and prepared the manuscript. MW, LS assisted in data processing, results evaluation, and interpretation. LT, HP provided scientific input. CG founded g.tec medical engineering GmbH and supervised the project.

FUNDING

This work was funded by g.tec medical engineering GmbH.

REFERENCES

Cai, X., Ding, C., Nie, F., and Huang, H. (2013). "On the Equivalent of Low-Rank Linear Regressions and Linear Discriminant Analysis Based Regressions," in Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. Editor I. Volosyak (New York, NY; London United Kingdom: Association for Computing Machinery; Computational Intelligence and Neuroscience). doi:10.1145/2487575.2487701

Chen, Z., Jin, J., Daly, I., Zuo, C., Wang, X., and Cichocki, A. (2020). Effects of Visual Attention on Tactile P300 Bci. *Comput. Intelligence Neurosci.* 2020, 1–11. doi:10.1155/2020/6549189

Comerchero, M. D., and Polich, J. (1999). P3a and P3b from Typical Auditory and Visual Stimuli. *Clin. Neurophysiol.* 110, 24–30. doi:10.1016/s0168-5597(98)00033-1

Donchin, E., and Smith, D. B. D. (1970). The Contingent Negative Variation and the Late Positive Wave of the Average Evoked Potential. *Electroencephalography Clin. Neurophysiol.* 29, 201–203. doi:10.1016/0013-4694(70)90124-0

Duncan-Johnson, C. C., and Donchin, E. (1977). On Quantifying surprise: The Variation of Event-Related Potentials with Subjective Probability. *Psychophysiology* 14, 456–467. doi:10.1111/j.1469-8986.1977.tb01312.x

Farwell, L. A., and Donchin, E. (1988). Talking off the Top of Your Head: toward a Mental Prosthesis Utilizing Event-Related Brain Potentials. *Electroencephalography Clin. Neurophysiol.* 70, 510–523. doi:10.1016/0013-4694(88)90149-6

Gao, S., Wang, Y., Gao, X., and Hong, B. (2014). Visual and Auditory Brain-Computer Interfaces. *IEEE Trans. Biomed. Eng.* 61, 1436–1447. doi:10.1109/TBME.2014.2300164

Groppe, D. M., Urbach, T. P., and Kutas, M. (2011). Mass Univariate Analysis of Event-Related Brain Potentials/fields I: A Critical Tutorial Review. *Psychophysiology* 48, 1711–1725. doi:10.1111/j.1469-8986.2011.01273.x

Gruenwald, J., Znobishchev, A., Kapeller, C., Kamada, K., Scharinger, J., and Guger, C. (2019). Time-variant Linear Discriminant Analysis Improves Hand Gesture and finger Movement Decoding for Invasive Brain-Computer Interfaces. *Front. Neurosci.* 13, 901. doi:10.3389/fnins.2019.00901

g.tec neurotechnology GmbH (2021). Unicorn Blondy Check. Available at: https://www.unicorn-bi.com/blondy-check/ (Accessed 01 27, 2021).

g.tec neurotechnology GmbH (2020). Unicorn Hybrid Black. Available at: https://www.unicorn-bi.com/brain-interface-technology/ (Accessed 01 27, 2021).

Guger, C., Daban, S., Sellers, E., Holzner, C., Krausz, G., Carabalona, R., et al. (2009). How many People Are Able to Control a P300-Based Brain–Computer Interface (Bci)? *Neurosci. Lett.* 462, 94–98. doi:10.1016/j.neulet.2009.06.045

Guger, C., Ortner, R., Dimov, S., and Allison, B. (2016). "A Comparison of Face Speller Approaches for P300 BCIs," in 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, Hungary, 9–12 Oct 2016, (IEEE), 004809–004812.

Guger, C., Spataro, R., Allison, B. Z., Heilinger, A., Ortner, R., Cho, W., et al. (2017). Complete Locked-In and Locked-In Patients: Command Following Assessment and Communication with Vibro-Tactile P300 and Motor Imagery Brain-Computer Interface Tools. *Front. Neurosci.* 11, 251. doi:10.3389/fnins.2017.00251

Krusienski, D., Sellers, E., McFarland, D., Vaughan, T., and Wolpaw, J. (2008). Toward Enhanced P300 Speller Performance. *J. Neurosci. Methods* 167, 15–21. doi:10.1016/j.jneumeth.2007.07.017

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique*. second edition. A Bradford Book. Cambridge: Mit Press. doi:10.1093/actrade/9780199672875.001.0001

Lugo, Z. R., Rodriguez, J., Lechner, A., Ortner, R., Gantner, I. S., Laureys, S., et al. (2014). A Vibrotactile P300-Based Brain–Computer Interface for Consciousness Detection and Communication. *Clin. EEG Neurosci.* 45, 14–21. doi:10.1177/1550059413505533

Magliero, A., Bashore, T. R., Coles, M. G., and Donchin, E. (1984). On the Dependence of P300 Latency on Stimulus Evaluation Processes. *Psychophysiology* 21, 171–186. doi:10.1111/j.1469-8986.1984.tb00201.x

Martens, S. M. M., Hill, N. J., Farquhar, J., and Schölkopf, B. (2009). Overlap and Refractory Effects in a Brain–Computer Interface Speller Based on the Visual P300 Event-Related Potential. *J. Neural Eng.* 6, 026003. doi:10.1088/1741-2560/6/2/026003

McCarthy, G., and Donchin, E. (1981). A Metric for Thought: a Comparison of P300 Latency and Reaction Time. *Science* 211, 77–80. doi:10.1126/science.7444452

Mugler, E. M., Ruf, C. A., Halder, S., Bensch, M., and Kubler, A. (2010). Design and Implementation of a P300-Based Brain-Computer Interface for Controlling an Internet Browser. *IEEE Trans. Neural Syst. Rehabil. Eng.* 18, 599–609. doi:10.1109/tnsre.2010.2068059

Nicolas-Alonso, L. F., and Gomez-Gil, J. (2012). Brain Computer Interfaces, a Review. *Sensors* 12, 1211–1279. doi:10.3390/s120201211

Pfurtscheller, G. (2001). Functional Brain Imaging Based on Erd/ers. *Vis. Res.* 41, 1257–1260. doi:10.1016/S0042-6989(00)00235-2

Polich, J., Ellerson, P. C., and Cohen, J. (1996). P300, Stimulus Intensity, Modality, and Probability. *Int. J. Psychophysiology* 23, 55–62. doi:10.1016/0167-8760(96)00028-1

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

M. D. Rugg and M. G. H. Coles (Editors) (1995). *Electrophysiology of Mind: Event-Related Brain Potentials and Cognition. Electrophysiology of Mind: Event-Related Brain Potentials and Cognition*. New York, NY, USPage: Oxford University Press, xii, 220.

Rugg, M. D., and Coles, M. G. H. (1996). *Electrophysiology of Mind*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780198524168.001.0001

Sato, M., Kodama, N., Sasaki, T., and Ohta, M. (1996). Olfactory Evoked Potentials: Experimental and Clinical Studies. *J. Neurosurg.* 85, 1122–1126. doi:10.3171/jns.1996.85.6.1122

Schomer, D. L., and Silva, F. L. D. (2010). *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. 6th revised edition edn. Philadelphia: Lippincott Williams & Wilki.

Sellers, E. W., and Donchin, E. (2006). A P300-Based Brain–Computer Interface: Initial Tests by Als Patients. *Clin. Neurophysiol.* 117, 538–548. doi:10.1016/j.clinph.2005.06.027

Serby, H., Yom-Tov, E., and Inbar, G. F. (2005). An Improved P300-Based Brain-Computer Interface. *IEEE Trans. Neural Syst. Rehabil. Eng.* 13, 89–98. doi:10.1109/TNSRE.2004.841878

Takano, K., Komatsu, T., Hata, N., Nakajima, Y., and Kansaku, K. (2009). Visual Stimuli for the P300 Brain–Computer Interface: A Comparison of white/gray and green/blue Flicker Matrices. *Clin. Neurophysiol.* 120, 1562–1566. doi:10.1016/j.clinph.2009.06.002

Thulasidas, M., Guan, C., and Wu, J. (2006). Robust Classification of Eeg Signal for Brain–Computer Interface. *IEEE Trans. Neural Syst. Rehabil. Eng.* 14, 24–29. doi:10.1109/TNSRE.2005.862695

Walter, W. G., Cooper, R., Aldridge, V. J., McCALLUM, W. C., and Winter, A. L. (1964). Contingent Negative Variation : An Electric Sign of Sensori-Motor Association and Expectancy in the Human Brain. *Nature* 203, 380–384. doi:10.1038/203380a0

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

Wolpaw, J. R. (2012). *Brain-computer Interfaces: Principles and Practice*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780195388855.001.0001

Woodman, G. F. (2010). A Brief Introduction to the Use of Event-Related Potentials in Studies of Perception and Attention. *Atten. Percept. Psychophys.* 72, 2031–2046. doi:10.3758/BF03196680

Zhang, Y., Zhao, Q., Jin, J., Wang, X., and Cichocki, A. (2012). A Novel Bci Based on Erp Components Sensitive to Configural Processing of Human Faces. *J. Neural Eng.* 9, 026018. doi:10.1088/1741-2560/9/2/026018

Conflict of Interest: NS, MW, LS, and CG were employed by g.tec Guger Technologies OG/g.tec medical engineering 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.

frontiers in Human Neuroscience

OPINION
published: 29 June 2021
doi: 10.3389/fnhum.2021.711407

Affective Brain-Computer Music Interfaces—Drivers and Implications

*Elisabeth Hildt**

Center for the Study of Ethics in the Professions, Illinois Institute of Technology, Chicago, IL, United States

Keywords: brain-computer interaction, music, direct to consumer, ethics, enhancement, privacy, affective brain-computer interface, neuroessentialism

Brain-computer interfaces (BCIs) allow users to control a computer or other device with their brain activity. While BCI technology has been developed and used primarily in medical contexts, a broad spectrum of non-clinical applications is on the horizon, including fields like concentration management, sleep improvement, music, and painting (Gürkök and Nijholt, 2013; Coates McCall and Wexler, 2020; Saha et al., 2021).

Some BCI applications directly translate brain activity to music performance, offering ways for people with physical disabilities as well as for artists to express their emotions through music (Eaton et al., 2015; Daly et al., 2016; Deuel et al., 2017; Williams and Miranda, 2018).

The focus of this article is on affective BCIs that allow to identify and influence a person's affective states. In addition to providing users with suggestions for the music they like, some affective brain computer music interface (aBCMI) applications aim at modulating the affective states of their users (Daly et al., 2016, 2020; Williams and Miranda, 2018; Ehrlich et al., 2019): Based on determining how listeners respond to certain types of music, music that influences their emotional states can be chosen. These affective BCIs detect correlates of a user's current affective state and attempt to modulate it by generating or selecting music that, for example, serves to increase happiness or reduce stress levels.

While the future development of this type of technology is largely unclear, for future non-clinical aBCMI home applications to be of interest to a broad spectrum of potential users, the technology not only has to prove attractive to a wide audience but also must be ethically sound. Against the background of recent developments in direct-to-consumer (DTC) devices, in what follows, I will discuss driving forces behind aBCMI technology development and social and ethical aspects of the technology, focusing on the perceived role of the brain, mood enhancement, and privacy-related aspects.

DIRECT-TO-CONSUMER DEVICES—RECENT DEVELOPMENTS

Based on research studies carried out with aBCMIs in complex research environments, several steps have been taken toward developing DTC home applications of aBCMI technology. For example, Kalaganis et al. (2016) developed a consumer BCI that serves to evaluate music in popular on-demand streaming services.

Among the existing wearable devices is the Mico system, a concept model developed in 2013 which allows individualized music choice[1]. Mico is short for "Music inspiration from your subconsciousness," it consists of headphones and an app for iPhone. According to the developers' homepage, the sensor in the headphone detects brainwaves; the mico app then analyzes the input and matches it to the closest "neural pattern" in the database to identify the user's "neural group." Based on this categorization, the system selects and plays music from a music database in line with the user's status.

[1] https://neurowear.com/mico/

OPEN ACCESS

Edited by:
Anton Nijholt,
University of Twente, Netherlands

Reviewed by:
Ian Daly,
University of Essex, United Kingdom
Duncan A. H. Williams,
University of Salford, United Kingdom

Correspondence:
Elisabeth Hildt
ehildt@iit.edu

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

Received: *18 May 2021*
Accepted: *02 June 2021*
Published: *29 June 2021*

Citation:
Hildt E (2021) Affective Brain-Computer Music Interfaces—Drivers and Implications. Front. Hum. Neurosci. 15:711407. doi: 10.3389/fnhum.2021.711407

The wearable EEG headset developed by Imec seeks to measure and influence the emotions of the person wearing the headset: According to the company[2], the system can learn the musical preferences of the users, compose, and play back music in real time that is in line with their preferences, and influence their emotions to achieve the desired emotional state.

The portable and wearable EEG device called Crown™ developed by Neurosity is advertised as being able to boost productivity[3]. According to the company's homepage, the system detects brainwaves and plays music from Spotify that helps users to focus. The increase in productivity brought about by the device seems to be quite small, however (Koetsier, 2021, cf. video interview).

While these are interesting developments, no DTC devices display research-grade results. It is currently unclear whether any of these devices can fulfill what they promise (see Coates McCall and Wexler, 2020). In view of this, it will be important that manufacturers and companies avoid exaggerated claims about the devices' presumed capabilities, not only for reasons of decency in selling the product but also to avoid users generating inadequate beliefs about their affective states. Unrealistic consumer expectations may result in a hype followed by disappointment (van Lente et al., 2013).

THE ROLE OF THE BRAIN

The aBCMI technology promises to provide an individualized music experience, based on the users' brain data. This individualized approach seeks to allow for easier access to music in line with an individual's preferences and needs. As advertised, the mico DTC device "frees the user from having to select songs and artists and allows users to encounter new music just by wearing the device[4]."

While this certainly alleviates the users from the burden of complex choices, it also narrows down individual decision-making. Instead of individual choice, users listen to what the technology suggests. While a user is free to accept or reject the choice made by the system, the system clearly influences the music someone listens to, be it through the music database provided or through the headset and data processing.

Generally speaking, aBCMI use raises awareness of the role of the brain and the relevance of brain-related data. Brain data serves to specify user affective states, categorize users into user types, and define the beginning and end points of affect-modulating interventions.

As Duncan Williams and Eduardo Miranda write in the context of music therapy (Williams and Miranda, 2018, p. 201):

> "The theoretical advantage of this approach over conventional music therapy approaches is that the BCMI is able to directly monitor the users' emotional state via physiological indices of emotion, which have the potential to be more robust and objective measures of emotion than user reports or even the expertise of the music therapist."

Accordingly, at least with research grade aBCMIs in the context of music therapy, the technology is expected to provide a more direct, more reliable measure of emotion than subjective reports, be it user self-assessments or reports by third persons.

It is worth mentioning that this position touches on tricky longstanding philosophical questions on the epistemic authority of claims about sensations (Baier, 1962): Who or what can be more reliable—a technology that depends on brain recordings or a person's first-person perspective that is based on introspection?

Instead of characterizing a person as being in a certain mood or liking certain music by watching their overt behavior or relying on their introspective reports, with this technology-based approach it is brainwaves that help to find out about a person's emotions, what music they like, and how their mood is influenced by music. By focusing not on a person and their behavior but on brain-related data, the role of the brain is stressed.

While it may be argued that this is just the mechanism upon which aBCMIs rely, when talking or writing about aBCMIs, it is important to avoid neuro-realist interpretations that consider brain activation patterns as the ultimate proof that a phenomenon is real and objective, as well as neuro-essentialist interpretations that see the brain as the self-defining essence of a person (see Racine et al., 2010; Reiner, 2011).

Phrases like "Mico-Music inspiration from your subconsciousness" or "This EEG headset can tell what music your brain likes" (Shankland, 2018), or the characterization of BCI-based music experiments as bringing "a new meaning to 'straight off the dome"' (Chung, 2017), all consider the brain to be the central actor, a substitute for the person.

MOOD ENHANCEMENT AND BRAIN AUGMENTATION

Even though the headsets of DTC devices resemble traditional headsets, and aBCMI-mediated music consumption shares considerable similarities with other forms of music consumption and automated playlist selection technologies, there is a significant difference in that aBCMIs aim at positively influencing affective states.

While the term enhancement—and the distinction between treatments and enhancements—has been used as a boundary concept to characterize interventions as outside the realm of medicine (Frankford, 1998; Juengst, 1998), in more general terms, enhancements are procedures to augment a person's physical or mental capabilities (Lebedev et al., 2018; Coates McCall and Wexler, 2020).

A number of enhancement approaches based on pharmaceuticals or neurotechnology including BCIs have been described in recent years (Zehr, 2015; Cinel et al., 2019). Non-clinical aBCMI home applications that seek to make healthy users feel happier or more relaxed can be considered mood enhancement technologies, devices that aim at helping users to get focused and increase their productivity can be seen in the context of cognitive enhancement.

[2]https://www.imec-int.com/en/articles/imec-and-holst-centre-introduce-eeg-headset-for-emotion-detection
[3]https://neurosity.co/
[4]https://neurowear.com/mico/

Whereas pharmacological mood enhancers such as Prozac or other selective serotonin reuptake inhibitors (SSRI) are prescription drugs used for purposes other than originally intended and bear the clear risk of negative systemic side effects (Schermer, 2015), the situation is different for aBCMIs. It is enhancement purposes that current DTC devices are developed for. The aBCMI technology is based on the individual user's brain data, and consists of an individualized approach which in principle could allow each user to navigate very fine grained affect modulations. While negative effects of the technology cannot be excluded, an EEG-based headset can be disconnected easily at any time.

Whereas positive effects of DTC aBCMI devices have not been proven yet, on the one hand, it may be argued that the technology seeks to enable authenticity and increase autonomy in that it allows users to maintain, influence, or attain their desired affective states. On the other hand, outsourcing the responsibility for regulating and controlling one's affective states can be seen as essentially inauthentic and as limiting one's capabilities and autonomy in that it increases dependence on technology (see Steinert and Friedrich, 2020).

PRIVACY AND DATA PROTECTION

In general, issues related to privacy play an important role whenever brain-related data is being collected and stored. In the context of neurotechnologies, several authors have stressed security and privacy risks and argued toward a right to mental privacy and a right to mental integrity (Ienca and Andorno, 2017; Ienca et al., 2018; Lavazza, 2018).

There are a number of privacy aspects to consider around aBCMIs, even though it seems questionable whether current aBCMIs can reveal any detailed information on a person's thoughts or preferences (Coates McCall and Wexler, 2020). In aBCMIs, each EEG recording gives some indication of a person's affective state at a certain point in time. Over a longer period, this will add up to a relatively detailed profile of a user's affective states. Data protection and privacy require that the individual user be in control of what is recorded, how the recordings are stored, and what is revealed and shared by the system about data analysis and classification results.

While the future of DTC aBCMIs is uncertain, with wearable, smartphone-compatible devices, privacy issues can be expected to become even more central. In general, brain-related data processed and stored on a smartphone connected to the internet is susceptible to being the subject of a multitude of data collection and sharing pathways. This could potentially include future individualized nudging or neuromarketing approaches (Ienca et al., 2018; Steinert and Friedrich, 2020).

Data protection measures (such as encryption of brain activity) will have to be established in order to prevent unauthorized access, sharing and use of brain-related data (Hernandez, 2016; Koetsier, 2021).

CONCLUSION

Despite these first steps toward developing non-clinical aBCMI home applications, the future of direct-to-consumer aBCMI technology is uncertain. A broad spectrum of challenges will have to be addressed, including electrode development and placement, user comfort, validity, reliability, privacy, and costs. To be attractive, the technology will have to provide some clear benefits to its users, be it regarding human performance, well-being, or leisure activities. At the same time, it will be important to avoid brain-centric interpretations and overreliance on technologically mediated affect management.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Baier, K. (1962). Smart on sensations. *Austral. J. Philos.* 40, 57–68. doi: 10.1080/00048406212341041

Chung, B. (2017). *Here's How to Make Music With Only Your MIND!* Available online at: https://www.vice.com/en/article/4xqv53/10-pieces-of-music-created-by-brainwaves

Cinel, C., Valeriani, D., and Poli, R. (2019). Neurotechnologies for human cognitive augmentation: current state of the art and future prospects. *Front. Human Neurosci.* 13:13. doi: 10.3389/fnhum.2019.00013

Coates McCall, I., and Wexler, A. (2020). "Peering into the mind? The ethics of consumer neuromonitoring devices," in *Ethical Dimensions of Commercial and DIY Neurotechnologies,* eds I. Bard and E. Hildt (Cambridge, MA: Elsevier, Academic Press), 1–22.

Daly, I., Nicolaou, N., Williams, D., Hwang, F., Kirke, A., Miranda, E., et al. (2020). Neural and physiological data from participants listening to affective music. *Sci. Data* 7:177. doi: 10.1038/s41597-020-0507-6

Daly, I., Williams, D., Kirke, A., Weaver, J., Malik, A., Hwang, F., et al. (2016). Affective brain-computer music interfacing. *J. Neural Eng.* 13:046022. doi: 10.1088/1741-2560/13/4/046022

Deuel, T. A., Pampin, J., Sundstrom, J., and Darvas, F. (2017). The encephalophone: a novel musical biofeedback device using conscious control of electroencephalogram (EEG). *Front. Hum. Neurosci.* 11:213. doi: 10.3389/fnhum.2017.00213

Eaton, J., Williams, D., and Miranda, E. (2015). The space between us: evaluating a multi-user affective brain–computer music interface. *Brain Computer Interfaces* 2, 103–116. doi: 10.1080/2326263X.2015.1101922

Ehrlich, S. K., Agres, K. R., Guan, C., and Cheng, G. (2019). A closed-loop, music-based brain-computer interface for emotion mediation. *PLoS ONE* 14:e0213516. doi: 10.1371/journal.pone.0213516

Frankford, D. M. (1998). "The treatment/enhancement distinction as an armament in the policy wars," in *Enhancing Human Traits. Ethical and Social Implications,* ed E. Parens (Washington, DC: Georgetown University Press), 70–94.

Gürkök, H., and Nijholt, A. (2013). "Affective brain-computer interfaces for arts," in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction* (Geneva), 827–831. doi: 10.1109/ACII.2013.155

Hernandez, A. (2016). *Brain Waves Technologies: Security in Mind? I Don't Think So, IOActive.* Available online at: https://ioactive.com/brain-waves-technologies-security-in-mind-i-dont-think-so/

Ienca, M., and Andorno, R. (2017). Towards new human rights in the age of neuroscience and neurotechnology. *Life Sci. Soc. Policy* 13:5. doi: 10.1186/s40504-017-0050-1

Ienca, M., Haselager, P., and Emanuel, E. J. (2018). Brain leaks and consumer neurotechnology. *Nat. Biotechnol.* 36, 805–810. doi: 10.1038/nbt.4240

Juengst, E. T. (1998). "What does enhancement mean?," in *Enhancing Human Traits. Ethical and Social Implications,* ed E. Parens (Washington, DC: Georgetown University Press), 29–47.

Kalaganis, F., Adamos, D. A., and Laskaris, N. (2016). "A consumer BCI for automated music evaluation within a popular on-demand music streaming service, "Taking Listener's Brainwaves to Extremes"," in *12th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI)* (Thessaloniki), 429–440.

Koetsier, J. (2021). *This 'Personal Brain Computer' Boosts Productivity by Sensing Your Brainwaves and Playing Music From Spotify*. Available online at: https://www.forbes.com/sites/johnkoetsier/2021/03/16/this-personal-brain-computer-boosts-productivity-for-900/?sh=339dd822a805

Lavazza, A. (2018). Freedom of thought and mental integrity: the moral requirements for any neural prosthesis. *Front. Neurosci.* 12:82. doi: 10.3389/fnins.2018.00082

Lebedev, M. A., Opris, I., and Casanova, M. F. (2018). Editorial: augmentation of brain function: facts, fiction and controversy. *Front. Syst. Neurosci.* 12:45. doi: 10.3389/fnsys.2018.00045

Racine, E., Waldman, S., Rosenberg, J., and Illes, J. (2010). Contemporary neuroscience in the media. *Soc Sci Med.* 71, 725–733. doi: 10.1016/j.socscimed.2010.05.017

Reiner, P. B. (2011). "The rise of neuroessentialism," in *The Oxford Handbook of Neuroethics,* eds J. Illes and B. J. Sahakian (New York, NY: Oxford University Press), 161–176.

Saha, S., Mamun, K. A., Ahmed, K., Mostafa, R., Naik, G. R., Darvishi, S., et al. (2021). Progress in brain computer interface: challenges and opportunities. *Front. Syst. Neurosci.* 15:578875. doi: 10.3389/fnsys.2021.578875

Schermer, M. (2015). "Ethics of pharmacological mood enhancement," in *Handbook of Neuroethics,* eds J. Clausen and N. Levy (Dordrecht: Springer), 1177–1190.

Shankland, S. (2018). *This EEG Headset Can Tell What Music Your Brain Likes*. Available online at: https://www.cnet.com/news/imec-eeg-headset-music-brain-waves/

Steinert, S., and Friedrich, O. (2020). Wired emotions: ethical issues of affective brain-computer interfaces. *Sci. Eng. Ethics* 26, 351–367. doi: 10.1007/s11948-019-00087-2

van Lente, H., Spitters, C., and Peine, A. (2013). Comparing technological hype cycles: towards a theory. *Technol. Forecast. Soc. Change* 80, 1615–1628. doi: 10.1016/j.techfore.2012.12.004

Williams, D. A. H., and Miranda, E. R. (2018). "BCI for music making. Then, now, and next," in *Brain–Computer Interfaces Handbook. Technological and Theoretical Advances,* eds C. S. Nam, A. Nijholt, and F. Lotte (New York, NY: CRC Press), 193–205.

Zehr, E. P. (2015). Future think: cautiously optimistic about brain augmentation using tissue engineering and machine interface. *Front. Syst. Neurosci.* 9:72. doi: 10.3389/fnsys.2015.00072

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

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 19 July 2021
doi: 10.3389/fnhum.2021.647839

An Open Source-Based BCI Application for Virtual World Tour and Its Usability Evaluation

Sanghum Woo[1], Jongmin Lee[2], Hyunji Kim[2], Sungwoo Chun[1], Daehyung Lee[1], Daeun Gwon[2] and Minkyu Ahn[1,2]*

[1] School of Computer Science and Electrical Engineering, Handong Global University, Pohang, South Korea, [2] Department of Information and Communication Engineering, Handong Global University, Pohang, South Korea

Brain–computer interfaces can provide a new communication channel and control functions to people with restricted movements. Recent studies have indicated the effectiveness of brain–computer interface (BCI) applications. Various types of applications have been introduced so far in this field, but the number of those available to the public is still insufficient. Thus, there is a need to expand the usability and accessibility of BCI applications. In this study, we introduce a BCI application for users to experience a virtual world tour. This software was built on three open-source environments and is publicly available through the GitHub repository. For a usability test, 10 healthy subjects participated in an electroencephalography (EEG) experiment and evaluated the system through a questionnaire. As a result, all the participants successfully played the BCI application with 96.6% accuracy with 20 blinks from two sessions and gave opinions on its usability (e.g., controllability, completeness, comfort, and enjoyment) through the questionnaire. We believe that this open-source BCI world tour system can be used in both research and entertainment settings and hopefully contribute to open science in the BCI field.

Keywords: P300, brain–computer interface, open-source application, serious game, usability

OPEN ACCESS

Edited by:
Anton Nijholt,
University of Twente, Netherlands

Reviewed by:
Angela Riccio,
Santa Lucia Foundation (IRCCS), Italy
Dong Ming,
Tianjin University, China

****Correspondence:***
Minkyu Ahn
minkyuahn@handong.edu

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

Received: *30 December 2020*
Accepted: *16 June 2021*
Published: *19 July 2021*

Citation:
Woo S, Lee J, Kim H, Chun S, Lee D, Gwon D and Ahn M (2021) An Open Source-Based BCI Application for Virtual World Tour and Its Usability Evaluation. Front. Hum. Neurosci. 15:647839. doi: 10.3389/fnhum.2021.647839

INTRODUCTION

Brain–computer interfaces are a form of technology that enables direct communication between humans and a computer through brain oscillation. Since it can improve the quality of life for disabled patients by providing a new communication channel, it has been given much attention and subsequently advanced over the last 40 years (Schmidt, 1980; Georgopoulos et al., 1986; Farwell and Donchin, 1988; Wolpaw et al., 2000; Curran and Stokes, 2003; Lotte et al., 2007; Nicolas-Alonso and Gomez-Gil, 2012; Hamedi et al., 2016; Abiri et al., 2019).

The P300 BCI is a paradigm popularly used in brain–computer interface (BCI) development (Fazel-Rezai et al., 2012). This paradigm uses the P300 component, which is a positive response raised about 300 msec after the presentation of an odd stimulus. Indeed, numerous studies have shown the feasibility of utilizing the P300 BCI with patients (e.g., patients with amyotrophic lateral sclerosis, ALS) and healthy subjects to communicate. For example, the P300 speller has been used as a tool to measure the performance of the P300 BCI system to see if the system can be used by ALS patients (Nijboer et al., 2008; Guy et al., 2018), to unveil the cognitive characteristics (e.g., temporal differences in visual stimulus processing compared with healthy people) of patients

(Riccio et al., 2018), or to confirm the efficacy of the system to many people (Guger et al., 2009). The P300 speller, a brainwave-based typewriter that uses the P300 BCI paradigm, usually consists of rows and columns with alphabetic/numeric characters and detects the intended character of the user based on the elicited P300 component by flashing rows/columns (Farwell and Donchin, 1988; Won et al., 2019). This system has several advantages. First, it shows a relatively high and stable performance (or information transfer rate), especially compared with motor imagery BCI (Guger et al., 2009; Cho et al., 2017; Won et al., 2019). The MI paradigm showed large variation in performances across subjects and users (Lee et al., 2019). Second, it provides an intuitive user interface (UI); what a user sees is what should be spelled. Third, it is designed for a communication purpose that meets the needs of patients (especially locked-in patients). Because of these advantages, the P300 speller has become a standard BCI application and has been used in investigating various research topics such as performance improvement (e.g., classification accuracy or information transfer rate) (Fazel-Rezai et al., 2012), the low-performance phenomenon called "BCI-illiteracy" (Carabalona, 2017; Won et al., 2019), calibration-less BCI (Lee et al., 2020), patient study (Guy et al., 2018; Velasco-Álvarez et al., 2019), and UI/UX in BCIs such as stimulation type (Guan et al., 2004), clustering of several characters (Fazel-Rezai and Abhari, 2009), 3D cubes (Qu et al., 2018), and facial based cues (Jin et al., 2012). Indeed, researchers have made great achievements and advancements with the P300 BCI speller. Moreover, considering that commercialized P300 BCI speller systems are in the market, it seems that the BCI application is already in the daily lives of people.

However, there are still issues to be considered for executing practical BCI applications. While BCI is often used with the disabled, the number of accessible applications is limited. Moreover, usability on the user side is sometimes overlooked in the research and development of BCIs. Usability is related to the ease and convenience of a given system to help the user achieve the desired goal and is also associated with an index of satisfaction (ISO 9241-11, 1998). Often, the available resources provided limit the user to a specific domain (Donchin et al., 2000). For example, the P300 speller is used as the standard for measuring the performance of the P300 BCI algorithm and signal-processing techniques. No matter how algorithms and signal processing techniques are developed, the end goal is for the application to effectively work for a specific purpose to meet the needs of users. Since the P300 speller is designed for typing characters, not playing games or surfing the internet, it is necessary to expand the available domain by developing new applications while simultaneously conducting research on suitable algorithms and signal processing techniques. Therefore, attention should also be paid to increasing the types of BCI applications and listening to the feedback of users while making great efforts to improve the performance of the BCI system (Ahn et al., 2014). Considering the limited mobility of potential BCI users, expanding the areas from communication to entertainment, hobbies, and daily work-related tasks is important.

Fortunately, recent studies have introduced various types of applications to the BCI field (see **Table 1**). Traditional targets (e.g., wheelchair and computer cursor) are often used for controls in research, but new BCI innovations are being researched, such as the exoskeleton (Frolov et al., 2017; Wang et al., 2018a), drone (Wang et al., 2018b), web browser (Zickler et al., 2011; Yu et al., 2012; Saboor et al., 2018), emailing (Zickler et al., 2011), and cleaning robot (Shao et al., 2020). In addition, the BCI field has produced more games, such as the traditional Tetris (Wang et al., 2019), action (Coyle et al., 2015) and games that stimulate rowing (Vourvopoulos et al., 2016), cart control (Wong et al., 2015), attention training (Rohani and Puthusserypady, 2015), as well as drawing (Botrel et al., 2015). In addition to the emergence of several applications, methods have been proposed to enhance usability and accessibility that should be considered for the development of BCIs for patients in terms of user-centered design (UCD) (Kübler et al., 2020).

Although the future of BCI looks very bright, an important complication hinders its progression. Over decades, applications of various themes have appeared, but these applications have not become widely accessible. To be exact, most BCI applications published in the literature are often closed (not shared) and documentations, such as user or developer manuals, are rarely created and provided. Thus, generally, these applications are not usable to other researchers.

From the point-of-view of the BCI researcher, this trend is fully understood, because application development is enormously expensive. In particular, the development of the P300 BCI application requires an extensive investment of time and effort for three main reasons. First, because it must operate online, the performance of the module responsible for data measurement and signal processing must be optimized for speed and accuracy. This is a common issue related to online application development. Second, because how well the P300 component is detected on the system determines the effectiveness of the application, it is necessary to search optimal parameters for stimulation (e.g., target-to-target interval, inter-stimulus interval, physical property, the distance between stimuli, and appropriate luminance of the stimulus for avoiding afterimage) under a given system design and apply it to the module in charge of the graphical user interface. Third, optimal bi-directional communication should be implemented to minimize the stimulus time lag and overall system delay that occur, as each module exchanges marker information. Because of costs, it is natural for a developer to accumulate results by conducting several studies using just his own application. However, when all developers do this, such large cost creates a high barrier for nonexperts, and the subsequent delay in research progress, consequently, may serve as a serious bottleneck that hinders the development of the BCI field. Therefore, just as developing the BCI application with new contents is important, sharing it with the research community is also crucial to expanding the field. We expect that diversifying application types will increase the efficiency of BCI research and ultimately contribute to leading the advancement of BCI.

So far, the obstacles that hinder the development of the current BCI have been mentioned, and methods to solve them have been

TABLE 1 | BCI application contents, paradigm, and platform.

Application contents	Article	Paradigm	Platform
Wheelchair control	Taher et al., 2015	EEG, Eye tracking	Emotive EPOC SDK, OpenViBE
	Yu et al., 2017	MI, P300	BCI 2000
	Bastos-Filho et al., 2018	SSVEP	C
Exoskeleton control	Frolov et al., 2017	MI	Matlab
	Wang et al., 2018a	SSVEP	-
Post-stroke rehabilitation using VR	Aamer et al., 2019	MI	Python, Unity 3D
Cursor control	Ma et al., 2017	MI, mVEP	-
Drone control	Wang et al., 2018b	SSVEP	Unreal Engine4, C++, Matlab
Web browser control	Zickler et al., 2011	P300	BCI 2000
	Yu et al., 2012	P300	Windows 32bit Platform Development Kit, Neuroscan
	Saboor et al., 2018	SSVEP	Microsoft VS C++
Emailing	Zickler et al., 2011	P300	BCI 2000
Cleaning robot	Shao et al., 2020	SSVEP	Matlab psychology toolbox, Bluetooth
Spelling	Lin et al., 2016	SSVEP, EMG	MATLAB
	Stawicki et al., 2017	SSVEP, Eye tracking	EyeTribe, Microsoft VS C++
IoT	Coogan and He, 2018	MI	Unity, BCI2000
Drawing game	Botrel et al., 2015	P300	BCI 2000
Action game	Coyle et al., 2015	MI	MATLAB Simulink
Cart control game	Wong et al., 2015	SSVEP	Microsoft VC++ 2010, DirectX SDK
Motion tracking game	Park et al., 2016	Neurofeedback	Unity 3D, Microsoft Kinect
Rowing game	Vourvopoulos et al., 2016	MI	Open ViBE, Unity, RehabNet Control Panel
Spatial navigation	Chen et al., 2017	SSVEP	Matlab
Tetris game	Wang et al., 2019	MI, SSVEP	Android SDK
VR: attention training	Rohani and Puthusserypady, 2015	P300	Microsoft Kinect, Unity 3D
	Ali and Puthusserypady, 2015	SSVEP	Unity, Adobe Photoshop, Autodesk 3DS Max
	Mercado et al., 2019	Neurofeedback	Unity, OpenViBE
VR: BCI system	McMahon and Schukat, 2018	MI	OpenViBE

MI, Motor imagery; SSVEP, steady state visual evoked potential; EMG, electromyogram; SDK, software development kit; VR, virtual reality.

suggested. Now is the time to take action on this. The aim of this study is not to propose a novel signal-processing algorithm or provide a consumer-grade application but instead introduce an open-source-based BCI application that can be easily reused and customized by BCI researchers at minimal costs (saving time, no need for platform charge). In this study, we developed a BCI world tour system (WTS) where a user can choose a touristic destination (country or city) and watch a movie that essentially takes them on a visual tour of the destination.

The P300 speller is appropriate for communication, but sometimes entertainment application is overlooked. Considering the limited mobility of the end users, providing various applications, such as entertainment, is important. Especially, it is unimaginable for them to travel in their limited circumstances. With this motivation, we chose virtual travel as the theme, which could help the end user to acquire travel experiences on their own, and contribute to enhance their self-efficacy, which is important for improving the quality of life (Bandura, 2010). Thus, we believe that the developed system could be meaningful for some end users (e.g., in the locked-in state) and also useful for other researchers. This application was built on three open source codes, and all the codes and detailed user manual are available in the Github repository (BCILab, 2020). Thus, anyone can access and use the application for their own purpose for free.

The following sections are organized as follows: In "Materials And Methods" section, we explain the development environment and scenario of the WTS as well as the experiment methods. The results from the questionnaire survey and performance from the online experiment are presented in "Results" section. Finally, further issues, such as the limitation of WTS, will be discussed in "Discussion" section.

MATERIALS AND METHODS

Application Development

Open Source Used

WTS operates through the interaction of three open source codes, which give us a competitive edge in terms of portability, scalability, online performance, and UI quality. They are OpenViBE, Python, and Unity 3D. Detailed information is as follows.

- OpenViBE for overall integration and scalability: an open-source software platform specialized for integrating various components of the BCI (Renard et al., 2010), OpenViBE

enables real-time acquisition, preprocessing, classification, and visualization of brain waves. The scenarios can be designed using function boxes, allowing users to design experiments more intuitively. Through this, portability was obtained in the process of collecting and processing the EEG signal and synchronizing it with the target application. Furthermore, OpenViBE is compatible with various EEG devices; thus, a device can be easily changed with minimal cost. However, there is also limitation. OpenViBE supports only Window or Linux operating systems; thus, it is hard to implement a BCI application running on mobile environment.
- Python for signal processing: a Python scripting box provided by OpenViBE was used in this system. In addition, scikit-learn, a state-of-the-art machine learning algorithms package, was used for signal processing and classification analysis (Pedregosa et al., 2011).
- Unity 3D for application: a game engine (https://unity3d.com) widely used in 3D game development, architectural visualizations, and real-time 3D animations. Under the integrated development and execution environment, developers can easily develop and debug applications. Since it supports multiple platforms, the application can be extended to various versions (Android, iOS, and personal computer).

Game Scenario and Contents

We wanted to give an indirect travel experience and provide control to the user. Thus, we designed the WTS to have options for the user to choose through the BCI and to provide an interesting content (e.g., video). In this sense, the WTS provides the names of countries or cities on the screen. For the purpose of the study, the destinations were chosen manually. However, the WTS is customizable, and the cities and contents can be changed by the developer or researchers for their own purpose.

Each step is limited to six commands to be the most suitable for human–computer interaction (HCI). Since cities are dependent on a specific continent, the system was designed with a region-based approach, so that users can more intuitively select the city they want. Therefore, we used a hybrid of the region-based paradigm and the single display paradigm that we mentioned earlier. Thirty-six target touristic places are selected and categorized into six continents, as shown in **Table 2**. The user interface was designed to have two steps. The first is choosing a continent and the second step is place selection, which is initiated right after the first step. To provide the information of the chosen place, we used short video clips available through the internet. The detailed list of videos is available in the WTS GitHub repository.

BCI Paradigm and Parameters

The WTS follows the conventional visual-evoked P300 BCI paradigm where target and non-target stimuli flicker in a randomized order (Squires et al., 1975; Katayama and Polich, 1996; Tarkka and Stokic, 1998; Strüber and Polich, 2002; Polich, 2007), while the BCI system processes the real-time EEG signal and detects the intended target.

A clearer P300 component is beneficial for maximum BCI performance, so it is necessary to set the optimal environment for this, namely the strength of the stimulus (e.g., brightness in the visual stimulus) and the time between the stimuli as well as the UI of the system to which the stimulus is given. In each step of the developed application, there are six stimuli—one target and five nontargets. This total is far smaller than the 36 in the conventional 6-by-6 P300 BCI speller, making the target-to-target interval (TTI) too short. This can be advantageous from a practical point of view by allowing the user to make quick selections. In addition, by adjusting the distance between adjacent commands, it is possible to classify targets in a shorter time, solving the problem of adjacencies being wrongfully detected as targets. However, reduction in average TTI may also lead to a smaller P300 amplitude (Fitzgerald and Picton, 1984; Polich, 1990; Gonsalvez and Polich, 2002) and may hinder the formation of prominent features of target epochs. This consequently causes degradation of performance in the BCI. Since the aim of this study is also to show the feasibility of the developed system, we simply used 20 for the number of blinks per stimulus to gain upper bound classification performance. Although the number of blinks in the WTS is greater than that of the P300 speller (normally 15 or fewer), the selection time for each step takes the same time as typing a character with the P300 speller. The inter-stimulus interval (ISI) is set to 187.5 ms (stimulus interval: 125 ms + blink time: 62.5 ms), which is the same as that of the conventional P300 Speller. However, the time for each selection is too long, making the system impractical. Thus, we performed offline analysis to obtain the optimal blink number, which we discuss in "Results" section.

TCP/IP Communication

For communication between OpenViBE to Unity 3D, TCP/IP was employed. OpenViBE provides a communication method called "TCP Tagging" that is reliable and gives the minimum overheads to the application (Foy, 2016). The WTS uses this protocol to send and receive messages between the OpenViBE and Unity 3D applications. We implemented the TCP/IP client code of Unity3d as concisely as possible to enable faster and more stable communication.

Application Evaluation

To evaluate the developed application, we conducted an EEG experiment with healthy participants. The BCI performance, EEG data, and opinions of users were collected for further analysis. This section describes the details of the experimental design and analysis procedure.

Participants

Ten healthy subjects participated in this experiment. Seven participants were female, and the average age of all the participants was 23.2 ± 1.72 years. The study was approved by the Public Institutional Bioethics Committee designated by the MOHW (P01-201812-11-004), and all the participants signed the consent form and were given information on the experiment and their rights before the experiment began.

TABLE 2 | Continents and touristic places used in the WTS.

Continents	Touristic places					
Europe	Paris	London	Rome	Barcelona	Iceland	Firenze
Asia	Seoul	Dubai	Hongkong	India	Tokyo	Shanghai
North America	Vancouver	New York	Las Vegas	Los Angeles	Chicago	Alaska
Oceania	Sydney	Melbourne	Fiji	New Zealand	Papua New Guinea	Vanuatu
South America	Barbados	Easter Island	Patagonia	Cusco	Rio de Janeiro	Buenos Aires
Africa	Egypt	Cape Town	Johannesburg	Nairobi	Pretoria East	Ethiopia

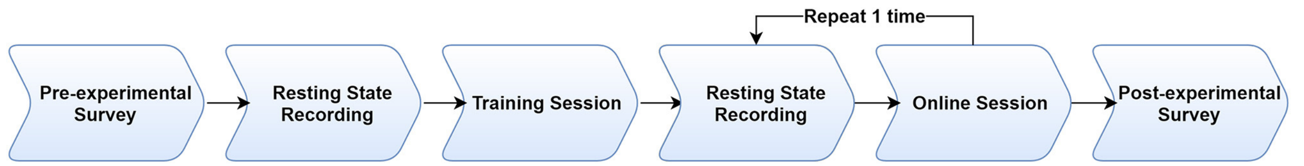


FIGURE 1 | Experimental procedure.

Experiment

Each experiment took about 50 min and consisted of a training session for generating the classifier and two subsequent online sessions where the subjects played the application with given targets. The subjects sat in front of a 27-inch LED monitor and were asked to follow the instructions. Each session started with a resting state recording block. This block consisted of open and closed eye conditions, each lasting 1 min. Both were conducted with relaxed bodies, and in the open eye condition, the subject was instructed to stare at the fixation cross on the screen. The training session consisted of presenting the subject with six buttons labeled with numbers 1–6, and each button was sequentially targeted and randomly flashed 30 times. This produced 180 target and 900 nontarget epochs. Based on the collected EEG signals, the classifier was constructed.

In the two subsequent online sessions, the subject played the application. The goal was to choose the instructed continent and touristic destination in order to watch its corresponding 10-s video. Each session consisted of six trials, and each trial started with the subject choosing first the target continent and then the target destination. The target continents and touristic places were randomly selected and provided to the subject as an instruction on the top of the screen. The only difference from the training session is that each button flashed 20 times. Once the place was selected, the video clip was played, and the next trial was initiated at the end of the video. Over two online sessions, the subject watched 12 movies of 12 touristic destinations, and this procedure produced 480 target and 2,400 nontarget epochs. The procedure of the experiment is further described in **Figure 1**. In both the training and online sessions, each subject was asked to look at the target stimulus and count the number of blinks. In addition, all sounds were muted, since unexpected or annoying sounds may distract the overall experiment.

Questionnaire

Pre- and post-experimental questionnaires were given to the subjects to evaluate the practical issues of the WTS from the perspective of the user. Questionnaire items were implemented in a Unity 3D environment to help each subject complete it easily and comfortably, and the results of the questionnaire were saved in an electronic text file for data analysis.

Since the aim of this evaluation is to collect the user feedback on how they accept this application, we designed naïve question items, which give us information about each part of the system. Thus, we did not construct any hypothesis. Basically, we referred to two published articles (Cho et al., 2017; Lee et al., 2019), and question items were organized according to a study (Cho et al., 2017) that collected BCI data from 52 subjects. Some items were adopted from the study, and we also added specific questions about the experience of the user with the application (e.g., Follow, Control, Enjoyment, and Completeness in **Table 3**). The items in each questionnaire are described below.

The pre-experiment questionnaire included questions focused on general information (e.g., history of neurological/mental disease, hours elapsed since smoking/drinking, hours slept the previous night), previous experience in a BCI experiment, and self-assessed scores of depression, mood, and expectation of the application in a 5-point Likert scale.

The question items in the post-experiment questionnaire were designed to assess application usability and gather opinions of the subjects. These questions ask the subjects to evaluate instructions of the experiment, controllability of the application, adequacy of playing time, and appropriateness of the surrounding environment. Finally, questions concerning the overall completeness of the WTS and enjoyment of the subject were asked to measure satisfaction. Additional details about the question-and-answer format of the pre/post questionnaires are listed in **Table 3**.

TABLE 3 | Items of pre/post questionnaires.

Questionnaires	Question items	Answer format
Pre	• Have you had brain or mental disease?	Yes or No
	• Have you ever participated in BCI experiment or game?	Yes or No
	• Write hours you slept the previous night.	1–24
	• Write hours elapsed since you had alcohol.	1–24, 0 if did not
	• Write hours elapsed since you had a cigarette.	1–24, 0 if did not
	• Evaluate your depression level. (Depression)	1–5 (Depressed)
	• Evaluate your mood level. (Mood)	1–5 (Excited)
	• Evaluate your expectation to BCI WTS. (Expectation)	1–5 (Interested)
Post	• Evaluate your mood level. (Mood)	1–5 (Excited)
	• Evaluate how well you followed the instruction. (Follow)	1–5 (Well)
	• Evaluate the controllability to operate the WTS. (Control)	1–5 (Easy)
	• Evaluate the playing time. (Length)	1–5 (Long)
	• Evaluate the comfort of surroundings. (Comfort)	1–5 (Comfortable)
	• Evaluate the completeness of the WTS. (Completeness)	1–5 (High)
	• Evaluate how much you enjoyed the WTS. (Enjoyment)	1–5 (Enjoyed)

Data Acquisition and Processing

For EEG acquisition, we used the Biosemi Active Two system (with 32 channels, 2,048 Hz sampling rate). During the experiment, these 32 electrodes were attached to the scalp of the subject according to the international standard 10–20 System (Jasper, 1958), and the brain signals were recorded from 32 locations (FP1, AF3, F7, F3, FC1, FC5, FC6, FC2, F4, F8, AF4, FP2, Fz, C3, CP1, CP5, CP6, CP2, C4, Cz, P7, P3, Pz, PO3, PO4, P4, P8, T7, T8, O1, Oz, and O2).

All EEG data acquired during the training session were used to construct a classifier that was used in the two subsequent online sessions. The procedure of the signal processing is presented in **Figure 2**.

First, the raw EEG was down-sampled from 2,048 to 512 Hz and re-referenced by the common average reference. This signal was spectrally (0.5–10 Hz) and temporally filtered (200–600 ms based on cue onset) to extract the only interesting section of the signal. Then, baseline correction and down-sampling to 128 Hz were performed. The amplitudes of each epoch over all 32 channels were converted into a long feature vector and the significant features were determined through the stepwise feature selection with the ordinary least square method ($p < 0.05$). In the training session, the selected amplitude features were used to train a linear classifier. In the online sessions, the same process was followed to produce a long-feature vector consisting of selected amplitudes overall time and channels. Then, this feature vector was fed into the constructed classifier in the training session. The classifier output for each blink has a hard label of 0 (nontarget) or 1 (target). All the outputs from the classifier across the blinks were summed per button, and a selection (place) with the highest value was chosen as a target. In this procedure, no artefact detection or rejection was performed; thus, all the epochs were used in the following analysis.

Analysis

Each subject played 24 selections (six continents and six place selections in each session) during the two online sessions. We counted the number of selections that were correctly classified through EEG and used the percentage value obtained by dividing the number of total selections as a final online performance. In each selection process, there were six stimuli and 20 epochs per stimulus, resulting in a total of 120 epochs (target: 20, nontarget: 100). The number of epochs per stimulus is tremendously important for the system response time. Thus, we also investigated accuracy by decreasing the different number of epochs (or blinks of each stimulus) per selection. To calculate the simulated accuracy, we first set N as the number of epochs that were used in the classification. Then N number of epochs were randomly selected from all the epochs in each selection and evaluated for target versus nontarget classification. This process was repeated 10 times and consequently yielded 10 accuracy estimates over the selection problems. Finally, the offline accuracy for N was calculated by averaging the 10 estimates. We calculated the offline accuracy with different Ns, which were 1, 5, 10, 15, and 20.

RESULTS

BCI World Tour System

All of the source codes and documents for the WTS can be found in the Github repository (BCILab, 2020). In addition, the repository includes the user manual of the application, so that any developer or researcher can easily modify and play the WTS for research or entertainment purposes. In the following section, we describe the developed application using state and system diagrams.

Figure 3 describes a state diagram of the developed application. The system starts with the initial state and the username is input. Then, the resting and training scenes are started. Once the training mode is completed, then the user can play through the play (online) mode. In the online mode, the map is positioned in the background, and the stimuli indicating the continents and touristic places are overlaid. To provide the new travel experience to the user, the background scene was designed to have touristic images (e.g., sky, airplane, world map, and tourist sites). However, during selection, the background changes to the same dark blue screen used in the training session.

The application viewed from the side of the developer is as follows: in the online mode, the user looks at the target stimulus to choose a continent while all the six stimuli randomly blink. Whenever a stimulus blinks, this moment is marked and transmitted to the Python module in OpenViBE. When the blinking period is done, the pre-trained stepwise linear discriminant analysis (SWLDA) algorithm from the training

FIGURE 2 | Procedure of the signal processing.

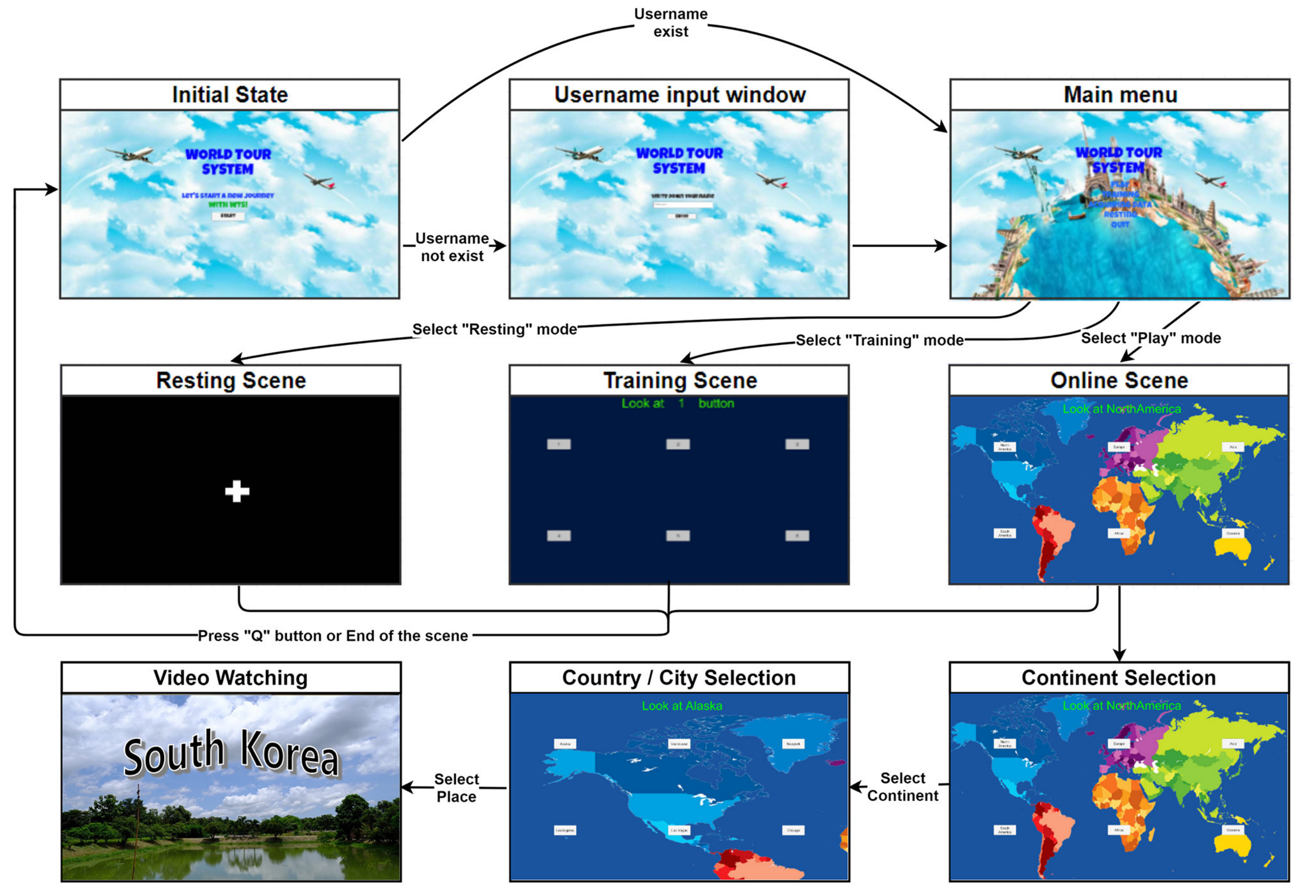


FIGURE 3 | The WTS state diagram.

mode classifies the given EEG signal and determines the continent. Based on the predicted continent, Unity 3D switches from the continental scene to the corresponding destination scene. Subsequently, the scene presents the new map of the chosen continent and the stimuli of six locations (country or city). Once a target place is determined by the same procedure used in continent selection, the corresponding video is played. The system diagram of the WTS is shown in **Figure 4**.

The WTS has the following features. First, it works with various EEG devices, because OpenViBE supports many different EEG devices. Second, customized algorithms can be used. OpenViBE provides a box "Python Scripting" that allows it to execute Python code (Bonnet, 2012). The box is used to process data entering, preprocessing, and leaving OpenViBE. This means that any algorithm implemented in Python can be reused in the WTS. Although SWLDA was used in this study for evaluation, developers can implement their own algorithm in Python script and use it for the main signal-processing code in the WTS. The sample code for signal processing is provided in the WTS repository. Third, video clips can be updated. Since the video source is independently managed, it can be replaced by longer, shorter, or even different multimedia sources. By updating the video sources, playing may provide different experiences.

Experimental Results

None of the subjects had a neurological or mental disease, and three of them (S1, S5, S6) have previous experiences with the P300 experiment. The mean sleeping time was 5.15 h per night. None of the subjects smoked a cigarette, and only a subject (S4) consumed alcohol 10 h before the experiment. In the following

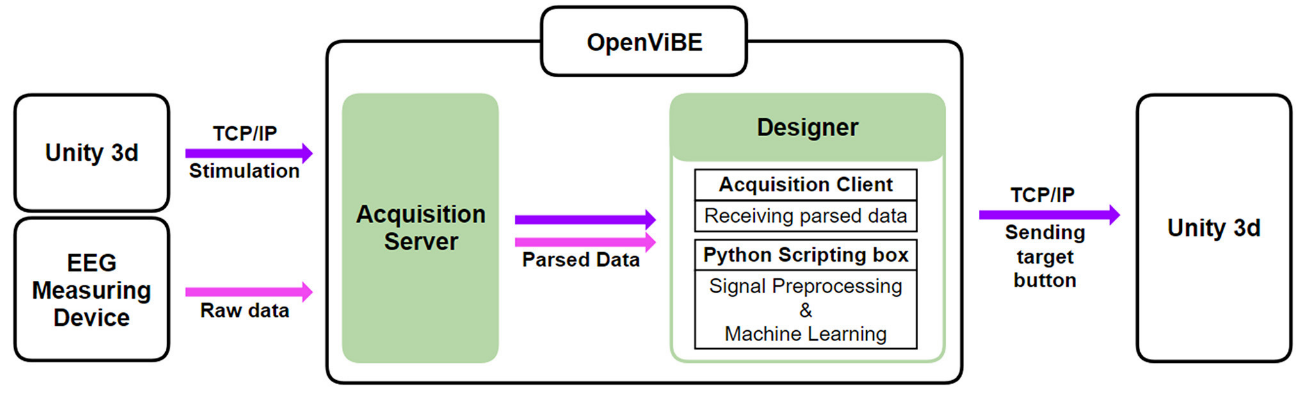


FIGURE 4 | The WTS system diagram.

TABLE 4 | Questionnaire results.

	Question items	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Mean(std)
Pre	Depression (1–5 Depressed)	2	3	3	4	3	3	1	1	2	1	2.3 ± 1.00
	Mood (1–5 Excited)	3	3	3	3	3	3	3	3	3	3	3.0 ± 0.00
	Expectation (1–5 Interested)	4	5	5	4	3	4	5	5	5	3	4.3 ± 0.78
Post	Mood (1–5 Excited)	3	2	3	1	2	2	3	3	3	3	2.5 ± 0.67
	Follow (1–5 Well)	5	3	4	5	5	3	5	3	5	4	4.2 ± 0.87
	Control (1–5 Easy)	5	5	3	5	5	5	5	3	5	5	4.6 ± 0.80
	Length (1–5 Long)	3	4	3	4	3	3	4	3	3	4	3.4 ± 0.49
	Comfort (1–5 Comfortable)	5	5	3	5	5	5	5	3	5	3	4.4 ± 0.92
	Completeness (1–5 Complete)	2	4	2	5	1	3	5	3	4	2	3.1 ± 1.30
	Enjoyment (1–5 Enjoyed)	3	4	1	4	2	3	4	4	3	2	3.0 ± 1.00

Note that a score of 3 means "Neutral."

section, the results from the questionnaire survey and online session are presented.

Questionnaire Results

The results of the questionnaires that were answered during the experiment are shown in **Table 4**. The subjects answered with an average score of 4.3 ± 0.78 (Expectation) for the pre-experimental question about the expectation of the WTS and an average score of 3 ± 1 (Enjoyment) for the post-experimental question about whether it was fun. The mood of the subjects before the experiment was close to neutral, showing an average score of 3, while it decreased to 2.5 ± 0.67 after the experiment. The subjects responded with average scores of 4.2 ± 0.87 for Follow and 4.6 ± 0.8 for Control, 3.1 ± 1.3 for Completeness, and 4.4 ± 0.92 for Comfort. When asked about the overall length of the application, they answered with an average score of 3.4 ± 0.49, which is slightly higher than 3 (Neutral). For more details, please refer to the discussion section.

Results From Online Experiment

Ten subjects successfully participated in one training and two online sessions. As we have mentioned previously, the EEG signals acquired during the training session were first analyzed, and then the classifier was constructed. **Figure 5A** is the picture of a representative subject in a prior pilot experiment. **Figure 5B** shows the target and nontarget ERP signals at the Cz channel averaged over epochs. Along with the Pz channel, the Cz channel is known for dominant occurrence of P3a (Johnson Jr, 1993). P3a is a subcomponent of P300 that occurs in the perceptual process when the P300 component is divided into perceptual and cognitive processes (Polich, 2007). To ensure that there is a significant difference between target and nontarget amplitudes in training data, the permutation test was performed (parametric two-sided t-test, alpha 0.05, 10,000 iterations) and false discovery rate (FDR) correction was performed (family-wise error rate = 0.05) for multiple testing correction (Benjamini and Yekutieli, 2005). As shown, significant clusters appeared in the ERP of all subjects except one (S8).

In the online sessions, the accuracy of each session was calculated. The subjects achieved a 95.8% average in the first session and 98.3% in the second session. The overall average accuracy was 96.6%. All subjects successfully played online sessions and eight subjects achieved 100%. The detailed accuracy for each subject is summarized in **Table 5**.

Offline Analysis

We conducted two offline analyses to check the significant channels and the influence of the number of blinks on BCI performance. The number of selected features during the training session varied across the subjects. Thus, to examine the

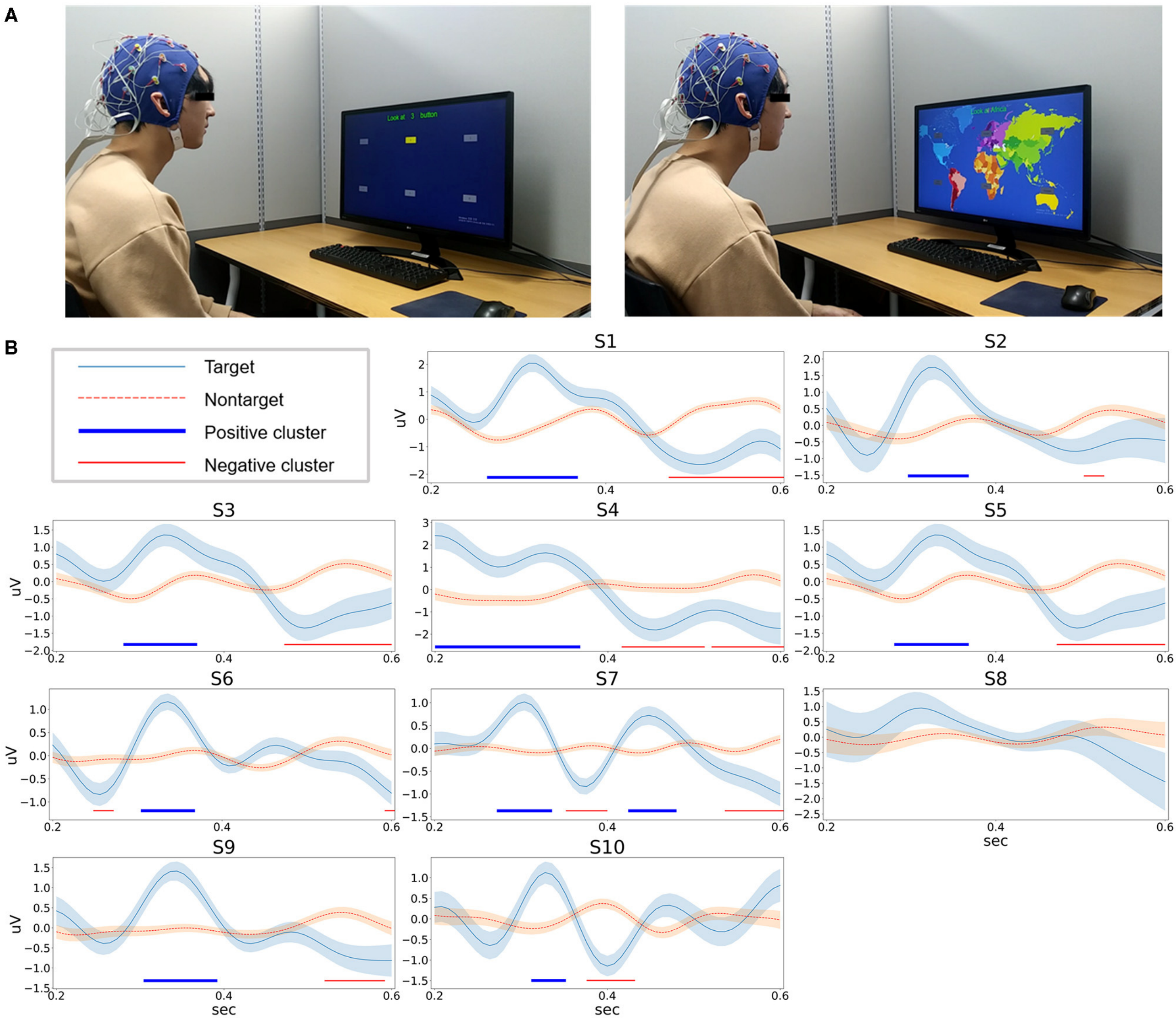


FIGURE 5 | (A) Pictures of training and online sessions from one representative subject in prior pilot experiment. **(B)** Averaged ERP signals at the Cz channel in the training session. Each plot shows the mean and standard error of the signal. In addition, positive and negative areas showing significant difference between target and nontarget epochs are shown at the bottom of each figure.

significant channels, we simply counted the number of selected features per channel. This procedure provides a histogram per subject. By summing up the histograms across all the subjects, we could obtain the result representing the degree of contribution to classification per channel. **Figure 6** represents the summed counts across all the subjects. As a result, an increasing tendency from frontal to occipital areas is observed. When checking the midline channels, this tendency becomes clearer (Fz $<$ Cz $<$ Pz $<$ Oz), which means that parieto-occipital channels are the main contributor in ERP classification.

An offline analysis was conducted to see if a smaller number of blinks per stimulus also work with reasonable accuracy. We checked the classification accuracy and information transfer rate (ITR) by changing the number of blinks from 1 to 20 (maximum). The result is shown in **Table 6**. It was revealed that the accuracy increases with a higher number of blinks, averaging 48.87, 76.5, 89.95, 93.7, and 96.66% for $N = 1, 5, 10, 15$, and 20, respectively. These increases were significant ($p < 0.05$, by Wilcoxon signed-rank test), but ITR peaks at $N = 5$ and the statistical test revealed that there is a significant difference ($p < 0.05$) between every pair except for $N = 5$ and $N = 10$ ($p > 0.05$).

DISCUSSION

In this study, we introduced an open-source BCI application, which uses the P300 BCI control paradigm. Through experiment and survey, we demonstrated the reasonable performance of this system and provided the opinion of the user. However,

TABLE 5 | Classification results from two online sessions.

Selection	S1		S2		S3		S4		S5		S6		S7		S8		S9		S10	
	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
1	O	O	O	O	O	O	O	O	X	O	O	O	O	O	O	O	O	O	O	O
2	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
3	O	O	O	O	O	O	O	O	X	O	O	O	O	O	O	O	O	O	O	X
4	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
5	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	X
6	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
7	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
8	O	O	O	O	O	O	O	O	O	X	O	O	O	O	O	O	O	O	O	O
9	O	O	O	O	O	O	O	O	X	O	O	O	O	O	O	O	O	O	O	O
10	O	O	O	O	O	O	O	O	X	O	O	O	O	O	O	O	O	O	O	O
11	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
12	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	X	O
Accuracy	100	100	100	100	100	100	100	100	66	91	100	100	100	100	100	100	100	100	91	83
ITR	5.09	5.09	5.09	5.09	5.09	5.09	5.09	5.09	1.71	3.82	5.09	5.09	5.09	5.09	5.09	5.09	5.09	5.09	3.82	3.01

Correctness is marked with O (correct) or X (incorrect) in each selection. Overall accuracies (%) and the corresponding ITR (bit/min) are presented in the last row.

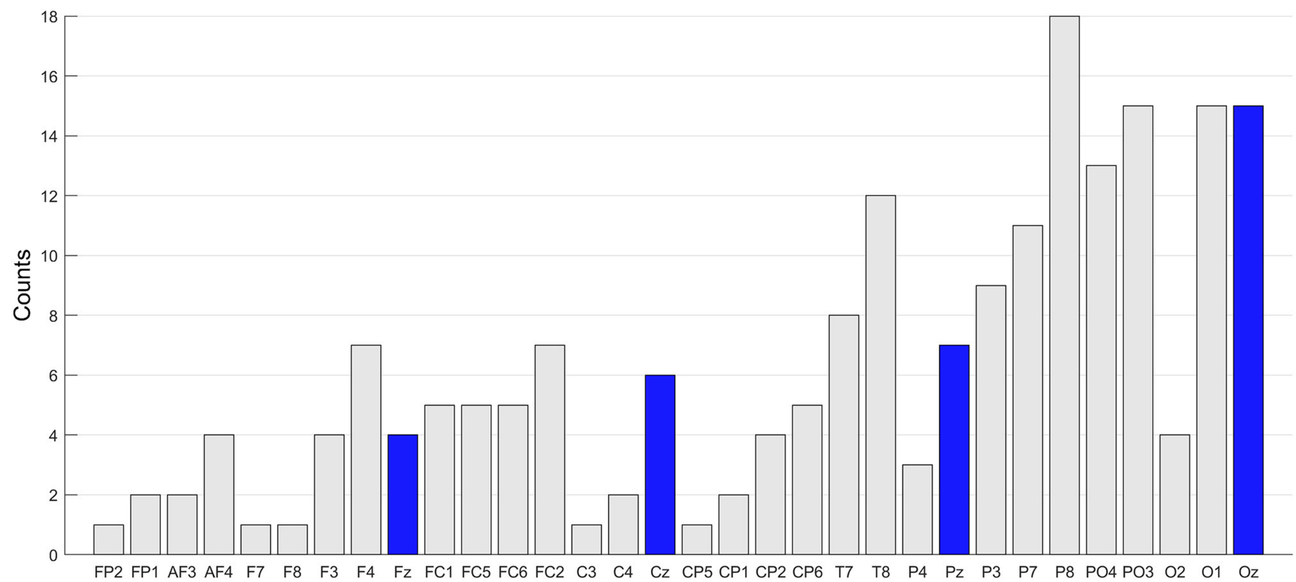


FIGURE 6 | Channel significance. The number of selected features per channel was summed across subjects. The channels are presented from frontal to occipital lobe for better visibility. The midline channels (Fz, Cz, Pz, and Oz) are marked with blue bar.

there are issues to discuss and limitations to the current version of the WTS. In the following subsection, we discuss several points observed in the results about the survey and online/offline analysis. Also, we present the potential limitations of this WTS and suggest future directions.

Questionnaire Study

Most BCI studies focus on system performance (e.g., classification accuracy), while the subjective opinion of BCI application is overlooked. However, because subjects are the potential users of BCI applications, their opinions are valuable to evaluate the overall usability of a BCI application and further improving the system. Some studies have used questionnaires to learn how users feel about BCI systems (Allison, 2009; Guger et al., 2009; Fazel-Rezai et al., 2012; Ahn et al., 2014, 2018). In this study, we also used questionnaires to collect personal information of subjects, system usability, and mood/enjoyment of users. Depending on the goal of the evaluation, the question items may vary, but we think that some general question items may be still useful in evaluating BCI applications. We suggest the following: (1) personal information (e.g., age, sex, BCI experience,

TABLE 6 | Accuracy results across a different number of blinks.

Number of blinks (Response time : ISI + system delay)		Accuracy/ITR										
		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Mean (std)
1 (6.62 s)	Acc.	49.16	50	42.50	66.25	36.25	50.00	52.50	69.58	41.66	30.83	**48.87 ± 11.50**
	ITR	3.66	3.84	2.41	7.97	1.45	3.84	4.39	8.99	2.27	0.79	**3.96 ± 2.65**
5 (11.37 s)	Acc.	83.75	70.83	71.66	93.75	52.08	83.75	86.25	95.85	81.25	45.83	**76.50 ± 15.76**
	ITR	8.27	5.47	5.63	11.10	2.50	8.27	8.91	11.82	7.67	1.75	**7.32 ± 3.32**
10 (17.75 s)	Acc.	95.00	89.16	88.33	100	69.58	98.33	99.16	99.16	90.41	70.41	**89.95 ± 10.80**
	ITR	7.38	6.21	6.06	8.74	3.35	8.19	8.44	8.44	6.44	3.45	**6.67±1.98**
15 (23.87 s)	Acc.	99.58	95	95	100	73.75	98.33	100	100	98.33	77.08	**93.70 ± 9.30**
	ITR	6.37	5.49	5.49	6.50	2.88	6.09	6.50	6.50	6.09	3.21	**5.51 ± 1.36**
20 (30.5 s)	Acc.	100	100	100	100	79.16	100	100	100	100	87.50	**96.66 ± 6.92**
	ITR	5.09	5.09	5.09	5.09	2.68	5.09	5.09	5.09	5.09	3.44	**4.68 ± 0.87**

Accuracy (and ITR) is presented over different number of blinks (N = 1–20) per subject. Mean and standard deviation (std) are presented in the last column.

disease history, and sleep hours); (2) system side (e.g., controllability, response time, overall completeness, UI/UX, and instruction); and (3) user side (e.g., mood, enjoyment, fear, difficulty, familiarity, expectation, and satisfaction). Perhaps, there may be more items, but we believe that considering these three categories together will help to better understand the opinions of users and ultimately further improve BCI applications.

Opinions of the subjects were obtained through the questionnaire items, and we can conclude the following based on the scores: Expectation is high, while Completeness and Enjoyment were not. As mentioned earlier, usability also includes helping a given system achieve the goals that users crave, so to optimize the usability of the system, it must contain what the user wants to achieve. The high expectation score supports that the WTS satisfies this condition. Thus, the WTS may need to be improved in UI/UX rather than system performance to increase user satisfaction. For example, the city video playback time was limited to 10 s for a smooth and short experiment and the content may not be satisfying to users. Therefore, it is necessary to improve the UI, video clips, button selection speed, etc. so that it can be more familiar to users.

Next, because Control and online accuracy are higher than Follow, it can be assumed that the WTS is effectively using the BCI system to reflect the intention of the user. Since the P300 epoch shown in **Figure 5** formed through the preprocessing process preserves the positive and negative components shown in the previous study (Polich, 2007), we think it has cleared the doubt of readers about the high system accuracy. Finally, for the question concerning the length of playing time, most of the subjects were not satisfied, sharing that they found the response time to be too long and somewhat boring. Therefore, offline analysis was performed to reduce the number of blinks; and in the next version, the reduced number of blinks can be used to shorten the system response time. However, the approach of the survey may be limited, since it was designed to measure simple opinion. Thus, some points might be missed. We think that the feedback of users is valuable information to update a BCI system. Also, certain guidelines for system design (Jeunet et al., 2018) or training protocol (Mladenović, 2021) would be considered from the initial phase of developing a new BCI application.

Improving Response Time

In the experiment, we used 20 blinks per stimulus, which led to a long response time—about 30.5 s for a selection. An offline analysis was performed to obtain a reasonable number of blinks. Ideally, the number should be small enough to shorten the response time but also yield good performance for use in a BCI. In **Table 6**, the average classification accuracy close to 90% is obtained at $N = 10$, and it yields 17.75 s for the response time for a selection in the WTS. On the other hand, ITR is relatively high at $N = 5$ and $N = 10$. Statistical test revealed that the two cases are not significantly different in ITR, but accuracy is statistically higher in $N = 10$ than in $N = 5$. Interestingly, six subjects already exceeded 90% at $N = 10$, and two subjects were close to 90%. Considering these results, we may choose $N = 10$, since it shows a relatively good ITR and high classification accuracy that is around 90%. Then, we can reduce the response time of the WTS by almost half. A more flexible approach rather than fixing the number of flashes can be used as introduced in Thomas et al. (2014) to efficiently running BCI with the aim of shortening the response time, or other control paradigms, such as steady state visual evoked potential (SSVEP), can be used for faster response time. However, visual fatigue should be considered before using it. SSVEP may cause more eye (or other modality) fatigue than the P300 because of persistent stimulation (Cao et al., 2014).

Improving Performance

There is another thing to note about the offline analysis results: The number of required blinks for good BCI performance seemed to vary across subjects. This may be related to the variation of ERP peaks across subjects (Won et al., 2019). In various studies, performance variation is one of the issues to be resolved (Guger et al., 2003, 2009, 2012; Ahn et al., 2013, 2018;

Ahn and Jun, 2015; Cho et al., 2015), so an in-depth analysis of this observation should be done to understand it in more detail. As noted in Data Acquisition and Processing section, no artefact rejection was performed in the system; thus, we believe that introducing a better machine learning technique or artefact rejection may help to improve the performance while decreasing the number of blinks for shorter response time and reduce the performance gap between subjects (Xiao et al., 2019).

User Adaptation

Table 5 presents the online accuracy of each session. Interestingly, a subject (S5) showed very different accuracies of 66% in the first and 91% in the second session. Since the SWLDA algorithm used in the WTS is not adaptively updated during online sessions, we interpret that the subject might adapt to the WTS. In other words, this result suggests that some users need time to get used to playing a certain BCI. However, the number of subjects who show this tendency and the required length should be investigated with more cases and UX issues in the BCI application. Furthermore, the standardized experimental protocols may be helpful for understanding or minimizing the performance variability among participants (Mladenović, 2021).

Limitations and Future Study

Although we demonstrated the applicability of the WTS, there are still limitations from a practical viewpoint. First, the number of commands that can be selected for each step is limited. Although there are only six continents, each continent has numerous cities. Therefore, we can increase the number of cities to choose from for each step. Since this system is open-source, it will be possible to increase commands for cities. However, as mentioned in Materials and Methods section, as the number of commands increases, the distance between adjacent commands becomes shorter, and an error in which they are misclassified as targets may occur. There are several studies that can help increase the number of commands (up to 100) while decreasing their size, so it is worth considering in future research (Xu et al., 2018, 2020).

Second, in the current version, the interaction between a user and the system is somewhat limited. There is no "move-back" or "pause" command. This means users should wait until the end of a selected video being played. In this sense, the system may be considered as not dynamical. Currently, the WTS is open to the public, thus touristic videos/names or command buttons can be changed for the purpose of the study by updating video files or source codes. However, the limitation of the interaction process in the current version should be considered before the actual use of the system and ultimately updated to provide better user-friendly UI/UX in the future.

Third, as a typical BCI application, the WTS also requires training time for generating a classifier to be used in the online session. However, this is one of the major obstacles hindering the progress of BCI applications. To be a more practical application, the training mode should be minimized or removed. Numerous studies are underway in the field to construct this general classifier (Kindermans et al., 2014a,b; Verhoeven et al., 2017; Eldeib et al., 2018; Lee et al., 2020). Usually, however, a general classifier requires a significant number of data samples, which can be achieved through transfer learning using data from one domain for another. Also, more complex machine learning algorithms (such as random forest, convolutional neural network, ensemble classifier) may be beneficial. In the future, we will also collect a large sample and investigate various models with the aim of achieving a calibration-less BCI application.

Fourth, the experiment was aimed at testing the system as a whole and performed with healthy subjects. We believe that the collected user feedback could be used in updating the system and this is also important. However, the system should be tested with the potential target group (e.g., patients) to understand the practical issues. This is beyond the scope of the present study, and we will consider this issue in future work. In addition, the current questionnaire was designed to simply confirm the opinion on the application using limited objective indicators. Thus, the result is somewhat limited in a sense like comparing with other BCI applications. A more systematic standard approach should be considered for system evaluation in the future (Lund, 2001).

Another limitation is that we only tested the WTS with a high-quality research purpose EEG device. However, considering that the BCI application should be easy enough for a naïve user to play with minimal knowledge and effort, the WTS should also be evaluated with devices with consumer-grade (cheap, easy, and possibly lesser channels) devices or dry electrodes.

CONCLUSIONS

We pointed out problems in the current BCI field and drew a big picture that may help the field to move forward. Also, we introduced a world tour system that is an open-source-based BCI application. The applicability of the WTS has been proven with an online experiment and questionnaire survey. All the codes and user manual for the WTS can be found in the GitHub repository. Thus, researchers and developers can easily use it for their own purposes because it comes with minimum costs (saving time, no need for platform charge). We hope that the arguments and the application will contribute to the BCI field, and ultimately, make many practical BCI applications emerge.

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 Public Institutional Bioethics Committee designated by the MOHW. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SW, SC, DL, and MA: conceptualization. SW and SC: methodology. SW, SC, DL, and HK: software. HK and

SW: questionnaire contents. SW: validation. SW and DG: investigation. JL: data curation. SW and MA: writing—original draft preparation, writing—review and editing, visualization, and project administration. MA: supervision. All the authors have read and approved the published version of the manuscript.

FUNDING

This research was funded by the National Research Foundation of Korea (NRF) (Grant No: 2019R1F1A1058844) and National Program for Excellence in Software at Handong Global University (Grant No: 2017-0-00130).

REFERENCES

Aamer, A., Esawy, A., Swelam, O., Nabil, T., Anwar, A., and Eldeib, A. (2019). "BCI integrated with VR for rehabilitation," in *2019 31st International Conference on Microelectronics (ICM)* (IEEE), 166–169.

Abiri, R., Borhani, S., Sellers, E. W., Jiang, Y., and Zhao, X. (2019). A comprehensive review of EEG-based brain–computer interface paradigms. *J. Neural Eng.* 16:011001. doi: 10.1088/1741-2552/aaf12e

Ahn, M., Ahn, S., Hong, J. H., Cho, H., Kim, K., Kim, B. S., et al. (2013). Gamma band activity associated with BCI performance: simultaneous MEG/EEG study. *Front. Hum. Neurosci.* 7:848. doi: 10.3389/fnhum.2013.00848

Ahn, M., Cho, H., Ahn, S., and Jun, S. C. (2018). User's self-prediction of performance in motor imagery brain–computer Interface. *Front. Hum. Neurosci.* 12:59. doi: 10.3389/fnhum.2018.00059

Ahn, M., and Jun, S. C. (2015). Performance variation in motor imagery brain–computer interface: a brief review. *J. Neurosci. Methods* 243, 103–110. doi: 10.1016/j.jneumeth.2015.01.033

Ahn, M., Lee, M., Choi, J., and Jun, S. C. (2014). A review of brain–computer interface games and an opinion survey from researchers, developers and users. *Sensors* 14, 14601–14633. doi: 10.3390/s140814601

Ali, A., and Puthusserypady, S. (2015). "A 3D learning playground for potential attention training in ADHD: a brain computer interface approach," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE), 67–70.

Allison, B. (2009). "The I of BCIs: next generation interfaces for brain–computer interface systems that adapt to individual users," in *International Conference on Human–Computer Interaction* (Springer), 558–568.

Bandura, A. (2010). "Self-efficacy," in *The Corsini Encyclopedia of Psychology* (NewYork: Wiley), 1–3.

Bastos-Filho, T., Floriano, A., Couto, E., and Godinez-Tello, R. J. (2018). "Towards a system to command a robotic wheelchair based on independent SSVEP–BCI," in *Smart Wheelchairs and Brain–Computer Interfaces* (Amsterdam: Elsevier) 369–379.

BCILab (2020). AhnBCILab/P300BCIWTS. Available online at: https://github.com/AhnBCILab/P300BCIWTS (accessed March 19, 2021).

Benjamini, Y., and Yekutieli, D. (2005). False discovery rate–adjusted multiple confidence intervals for selected parameters. *J. Am. Stat. Assoc.* 100, 71–81. doi: 10.1198/016214504000001907

Bonnet, L. (2012). Tutorial—Level 2—Using Python with OpenViBE. OpenViBE. Available online at: http://openvibe.inria.fr/tutorial-using-python-with-openvibe/ (accessed January 24, 2020).

Botrel, L., Holz, E. M., and Kübler, A. (2015). Brain Painting V2: evaluation of P300-based brain–computer interface for creative expression by an end-user following the user-centered design. *Brain–Comput. Interf.* 2, 135–149. doi: 10.1080/2326263X.2015.1100038

Cao, T., Wan, F., Wong, C. M., da Cruz, J. N., and Hu, Y. (2014). Objective evaluation of fatigue by EEG spectral analysis in steady-state visual evoked potential-based brain–computer interfaces. *Biomed. Eng. Online* 13, 1–13. doi: 10.1186/1475-925X-13-28

Carabalona, R. (2017). The role of the interplay between stimulus type and timing in explaining BCI-illiteracy for visual P300-based brain–computer interfaces. *Front. Neurosci.* 11:363. doi: 10.3389/fnins.2017.00363

Chen, J., Zhang, D., Engel, A. K., Gong, Q., and Maye, A. (2017). Application of a single-flicker online SSVEP BCI for spatial navigation. *PLoS ONE* 12:e0178385. doi: 10.1371/journal.pone.0178385

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

Cho, H., Ahn, M., Kim, K., and Jun, S. C. (2015). Increasing session-to-session transfer in a brain–computer interface with on-site background noise acquisition. *J. Neural Eng.* 12:066009. doi: 10.1088/1741-2560/12/6/066009

Coogan, C. G., and He, B. (2018). Brain–computer interface control in a virtual reality environment and applications for the internet of things. *IEEE Access* 6, 10840–10849. doi: 10.1109/ACCESS.2018.2809453

Coyle, D., Stow, J., McCreadie, K., Li, C., Garcia, J., McElligott, J., et al. (2015). "Action games, motor imagery, and control strategies: toward a multi-button controller," in *Handbook of Digital Games and Entertainment Technologies*, 1–34.

Curran, E. A., and Stokes, M. J. (2003). Learning to control brain activity: A review of the production and control of EEG components for driving brain–computer interface (BCI) systems. *Brain Cogn.* 51, 326–336. doi: 10.1016/S0278-2626(03)00036-8

Donchin, E., Spencer, K. M., and Wijesinghe, R. (2000). The mental prosthesis: assessing the speed of a P300-based brain–computer interface. *IEEE Trans. Rehabil. Eng.* 8, 174–179. doi: 10.1109/86.847808

Eldeib, A. M., Sarhan, O., and Wahed, M. A. (2018). "Zero training processing technique for P300-based brain–computer interface" in *2018 IEEE 4th Middle East Conference on Biomedical Engineering (MECBME)* (IEEE), 112–115.

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

Fazel-Rezai, R., and Abhari, K. (2009). A region-based P300 speller for brain–computer interface. *Can. J. Electr. Comput. Eng.* 34, 81–85. doi: 10.1109/CJECE.2009.5443854

Fazel-Rezai, R., Allison, B. Z., Guger, C., Sellers, E. W., Kleih, S. C., and Kübler, A. (2012). P300 brain computer interface: current challenges and emerging trends. *Front. Neuroeng.* 5:14. doi: 10.3389/fneng.2012.00014

Fitzgerald, P. G., and Picton, T. W. (1984). The effects of probability and discriminability on the evoked potentials to unpredictable stimuli. *Ann. N. Y. Acad. Sci.* 425, 199–203. doi: 10.1111/j.1749-6632.1984.tb23533.x

Foy, N. (2016). Extensions: TCP Tagging (Software Tagging). *OpenViBE*. Available online at: http://openvibe.inria.fr/tcp-tagging/ (accessed January 24, 2020).

Frolov, A. A., Mokienko, O., Lyukmanov, R., Biryukova, E., Kotov, S., Turbina, L., et al. (2017). Post-stroke rehabilitation training with a motor-imagery-based brain–computer interface (BCI)-controlled hand exoskeleton: a randomized controlled multicenter trial. *Front. Neurosci.* 11:400. doi: 10.3389/fnins.2017.00400

Georgopoulos, A. P., Schwartz, A. B., and Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science* 233, 1416–1419. doi: 10.1126/science.3749885

Gonsalvez, C. J., and Polich, J. (2002). P300 amplitude is determined by target-to-target interval. *Psychophysiology* 39, 388–396. doi: 10.1017/S0048577201393137

Guan, C., Thulasidas, M., and Wu, J. (2004). "High performance P300 speller for brain–computer interface," in *IEEE International Workshop on Biomedical Circuits and Systems 2004*, S3–S5.

Guger, C., Allison, B. Z., Großwindhager, B., Prückl, R., Hintermüller, C., Kapeller, C., et al. (2012). How many people could use an SSVEP BCI? *Front. Neurosci.* 6:169. doi: 10.3389/fnins.2012.00169

Guger, C., Daban, S., Sellers, E., Holzner, C., Krausz, G., Carabalona, R., et al. (2009). How many people are able to control a P300-based brain–computer interface (BCI)? *Neurosci. Lett.* 462, 94–98. doi: 10.1016/j.neulet.2009.06.045

Guger, C., Edlinger, G., Harkam, W., Niedermayer, I., and Pfurtscheller, G. (2003). How many people are able to operate an EEG-based brain–computer interface (BCI)? *IEEE Trans. Neural Syst. Rehabil. Eng.* 11, 145–147. doi: 10.1109/TNSRE.2003.814481

Guy, V., Soriani, M.-H., Bruno, M., Papadopoulo, T., Desnuelle, C., and Clerc, M. (2018). Brain computer interface with the P300 speller: usability for disabled people with amyotrophic lateral sclerosis. *Ann. Phys. Rehabil. Med.* 61, 5–11. doi: 10.1016/j.rehab.2017.09.004

Hamedi, M., Salleh, S.-H., and Noor, A. M. (2016). Electroencephalographic motor imagery brain connectivity analysis for BCI: a review. *Neural Comput.* 28, 999–1041. doi: 10.1162/NECO_a_00838

ISO 9241-11 (1998). *Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs). Part 11: Guidance on Usability*. Geneva: International Organization for Standardization.

Jasper, H. H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalogr. Clin. Neurophysiol.* 10, 370–375.

Jeunet, C., Debener, S., Lotte, F., Mattout, J., Scherer, R., and Zich, C. (2018). *Mind the traps! Design guidelines for rigorous BCI experiments*. New York: CRC Press. doi: 10.1201/9781351231954-32

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

Johnson Jr, R. A. Y. (1993). On the neural generators of the P300 component of the event-related potential. *Psychophysiology* 30, 90–97. doi: 10.1111/j.1469-8986.1993.tb03208.x

Katayama, J., and Polich, J. (1996). P300 from one-, two-, and three-stimulus auditory paradigms. *Int. J. Psychophysiol.* 23, 33–40. doi: 10.1016/0167-8760(96)00030-X

Kindermans, P.-J., Schreuder, M., Schrauwen, B., Müller, K.-R., and Tangermann, M. (2014a). True zero-training brain–computer interfacing–an online study. *PLoS One* 9:e102504. doi: 10.1371/journal.pone.0102504

Kindermans, P.-J., Tangermann, M., Müller, K.-R., and Schrauwen, B. (2014b). Integrating dynamic stopping, transfer learning and language models in an adaptive zero-training ERP speller. *J. Neural Eng.* 11:035005. doi: 10.1088/1741-2560/11/3/035005

Kübler, A., Nijboer, F., and Kleih, S. (2020). "Hearing the needs of clinical users," in *Handbook of clinical neurology* (Amsterdam: Elsevier), 353–368.

Lee, J., Won, K., Kwon, M., Jun, S. C., and Ahn, M. (2020). CNN with large data achieves true zero-training in online P300 brain–computer interface. *IEEE Access* 8, 74385–74400. doi: 10.1109/ACCESS.2020.2988057

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

Lin, K., Cinetto, A., Wang, Y., Chen, X., Gao, S., and Gao, X. (2016). An online hybrid BCI system based on SSVEP and EMG. *J. Neural Eng.* 13:026020. doi: 10.1088/1741-2560/13/2/026020

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., and Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain–computer interfaces. *J. Neural Eng.* 4:R1. doi: 10.1088/1741-2560/4/2/R01

Lund, A. M. (2001). Measuring usability with the use questionnaire. *Usability Interf.* 8, 3–6.

Ma, T., Li, H., Deng, L., Yang, H., Lv, X., Li, P., et al. (2017). The hybrid BCI system for movement control by combining motor imagery and moving onset visual evoked potential. *J. Neural Eng.* 14:026015. doi: 10.1088/1741-2552/aa5d5f

McMahon, M., and Schukat, M. (2018). "A low-cost, open-source, BCI-VR prototype for real-time signal processing of EEG to manipulate 3D VR objects as a form of neurofeedback," in *2018 29th Irish Signals and Systems Conference (ISSC)* (IEEE), 1–6.

Mercado, J., Espinosa-Curiel, I., Escobedo, L., and Tentori, M. (2019). Developing and evaluating a BCI video game for neurofeedback training: the case of autism. *Multimed. Tools Appl.* 78, 13675–13712. doi: 10.1007/s11042-018-6916-2

Mladenović, J. (2021). Standardization of protocol design for user training in EEG-based brain–computer interface. *J. Neural Eng.* 18:011003. doi: 10.1088/1741-2552/abcc7d

Nicolas-Alonso, L. F., and Gomez-Gil, J. (2012). Brain computer interfaces, a review. *Sensors* 12, 1211–1279. doi: 10.3390/s120201211

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

Park, K., Kihl, T., Park, S., Kim, M.-J., and Chang, J. (2016). "Narratives and sensor driven cognitive behavior training game platform," in *2016 IEEE 14th International Conference on Software Engineering Research, Management and Applications (SERA)* (IEEE), 125–131. doi: 10.1109/SERA.2016.7516137

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830. doi: 10.5555/1953048.2078195

Polich, J. (1990). P300, probability, and interstimulus interval. *Psychophysiology* 27, 396–403. doi: 10.1111/j.1469-8986.1990.tb02333.x

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

Qu, J., Wang, F., Xia, Z., Yu, T., Xiao, J., Yu, Z., et al. (2018). A novel three-dimensional P300 speller based on stereo visual stimuli. *IEEE Trans. Hum. Mach. Syst.* 48, 392–399. doi: 10.1109/THMS.2018.2799525

Renard, Y., Lotte, F., Gibert, G., Congedo, M., Maby, E., Delannoy, V., et al. (2010). Openvibe: An open-source software platform to design, test, and use brain–computer interfaces in real and virtual environments. *Presence Teleoper. Virtual Environ.* 19, 35–53. doi: 10.1162/pres.19.1.35

Riccio, A., Schettini, F., Simione, 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

Rohani, D. A., and Puthusserypady, S. (2015). BCI inside a virtual reality classroom: a potential training tool for attention. *EPJ Nonlinear Biomed. Phys.* 3, 12. doi: 10.1140/epjnbp/s40366-015-0027-z

Saboor, A., Gembler, F., Benda, M., Stawicki, P., Rezeika, A., Grichnik, R., et al. (2018). "A browser-driven SSVEP-based BCI web speller," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (IEEE), 625–630.

Schmidt, E. M. (1980). Single neuron recording from motor cortex as a possible source of signals for control of external devices. *Ann. Biomed. Eng.* 8, 339–349. doi: 10.1007/BF02363437

Shao, L., Zhang, L., Belkacem, A. N., Zhang, Y., Chen, X., Li, J., et al. (2020). EEG-controlled wall-crawling cleaning robot using SSVEP-based brain–computer interface. *J. Healthc. Eng.* 2020:6968713. doi: 10.1155/2020/6968713

Squires, N. K., Squires, K. C., and Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalogr. Clin. Neurophysiol.* 38, 387–401. doi: 10.1016/0013-4694(75)90263-1

Stawicki, P., Gembler, F., Rezeika, A., and Volosyak, I. (2017). A novel hybrid mental spelling application based on eye tracking and SSVEP-based BCI. *Brain Sci.* 7:35. doi: 10.3390/brainsci7040035

Strüber, D., and Polich, J. (2002). P300 and slow wave from oddball and single-stimulus visual tasks: inter-stimulus interval effects. *Int. J. Psychophysiol.* 45, 187–196. doi: 10.1016/S0167-8760(02)00071-5

Taher, F. B., Amor, N. B., and Jallouli, M. (2015). "A multimodal wheelchair control system based on EEG signals and Eye tracking fusion," in *2015 International Symposium on Innovations in Intelligent SysTems and Applications (INISTA)* (IEEE), 1–8.

Tarkka, I. M., and Stokic, D. S. (1998). Source localization of P300 from oddball, single stimulus, and omitted-stimulus paradigms. *Brain Topogr.* 11, 141–151. doi: 10.1023/A:1022258606418

Thomas, E., Dauc,é, E., Devlaminck, D., Mah,é, L., Carpentier, A., Munos, R., et al. (2014). "CoAdapt P300 speller: optimized flashing sequences and online learning," in *6th International Brain Computer Interface Conference*.

Velasco-Álvarez, F., Sancha-Ros, S., García-Garaluz, E., Fernández-Rodríguez, Á., Medina-Juli,á, M. T., and Ron-Angevin, R. (2019). UMA-BCI speller: an easily configurable P300 speller tool for end users. *Comput. Methods Programs Biomed.* 172, 127–138. doi: 10.1016/j.cmpb.2019.02.015

Verhoeven, T., Hübner, D., Tangermann, M., Müller, K.-R., Dambre, J., and Kindermans, P.-J. (2017). Improving zero-training brain–computer interfaces by mixing model estimators. *J. Neural Eng.* 14:036021. doi: 10.1088/1741-2552/aa6639

Vourvopoulos, A., Ferreira, A., and i Badia, S. B. (2016). "NeuRow: an immersive VR environment for motor-imagery training with the use of brain–computer interfaces and vibrotactile feedback," in *International Conference on Physiological Computing Systems (SCITEPRESS)*, 43–53.

Wang, C., Wu, X., Wang, Z., and Ma, Y. (2018a). Implementation of a brain–computer interface on a lower-limb exoskeleton. *IEEE Access* 6, 38524–38534. doi: 10.1109/ACCESS.2018.2853628

Wang, M., Li, R., Zhang, R., Li, G., and Zhang, D. (2018b). A wearable SSVEP-based BCI system for quadcopter control using head-mounted device. *IEEE Access* 6, 26789–26798. doi: 10.1109/ACCESS.2018.2825378

Wang, Z., Yu, Y., Xu, M., Liu, Y., Yin, E., and Zhou, Z. (2019). Towards a hybrid BCI gaming paradigm based on motor imagery and SSVEP. *Int. J. Hum. Comput. Interact.* 35, 197–205. doi: 10.1080/10447318.2018.1445068

Wolpaw, J. R., Birbaumer, N., Heetderks, W. J., McFarland, D. J., Peckham, P. H., Schalk, G., et al. (2000). Brain–computer interface technology: a review of the first international meeting. *IEEE Trans. Rehabil. Eng.* 8, 164–173. doi: 10.1109/TRE.2000.847807

Won, K., Kwon, M., Jang, S., Ahn, M., and Jun, S. C. (2019). P300 Speller Performance Predictor Based on RSVP Multi-feature. *Front. Hum. Neurosci.* 13:261. doi: 10.3389/fnhum.2019.00261

Wong, C. M., Tang, Q., da Cruz, J. N., and Wan, F. (2015). "A multi-channel SSVEP-based BCI for computer games with analogue control," in *2015 IEEE international conference on computational intelligence and virtual environments for measurement systems and applications (CIVEMSA)* (IEEE), 1–6.

Xiao, X., Xu, M., Jin, J., Wang, Y., Jung, T.-P., and Ming, D. (2019). Discriminative canonical pattern matching for single-trial classification of ERP components. *IEEE Trans. Biomed. Eng.* 67, 2266–2275. doi: 10.1109/TBME.2019.2958641

Xu, M., Han, J., Wang, Y., Jung, T.-P., and Ming, D. (2020). Implementing over 100 command codes for a high-speed hybrid brain–computer interface using concurrent P300 and SSVEP features. *IEEE Trans. Biomed. Eng.* 67, 3073–3082. doi: 10.1109/TBME.2020.2975614

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

Yu, T., Li, Y., Long, J., and Gu, Z. (2012). Surfing the internet with a BCI mouse. *J. Neural Eng.* 9, 036012. doi: 10.1088/1741-2560/9/3/036012

Yu, Y., Zhou, Z., Liu, Y., Jiang, J., Yin, E., Zhang, N., et al. (2017). Self-paced operation of a wheelchair based on a hybrid brain–computer interface combining motor imagery and P300 potential. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 2516–2526. doi: 10.1109/TNSRE.2017.2766365

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.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 08 October 2021
doi: 10.3389/fnhum.2021.663049

Study of Auditory Brain Cognition Laws-Based Recognition Method of Automobile Sound Quality

Liping Xie[1,2], *Chihua Lu*[1,2], *Zhien Liu*[1,2]*, *Lirong Yan*[1,2] *and Tao Xu*[1,2]

[1] *Hubei Key Laboratory of Advanced Technology for Automotive Components, Wuhan University of Technology, Wuhan, China,* [2] *Foshan Xianhu Laboratory of the Advanced Energy Science and Technology Guangdong Laboratory, Foshan, China*

The research shows that subjective feelings of people, such as emotions and fatigue, can be objectively reflected by electroencephalography (EEG) physiological signals Thus, an evaluation method based on EEG, which is used to explore auditory brain cognition laws, is introduced in this study. The brain cognition laws are summarized by analyzing the EEG power topographic map under the stimulation of three kinds of automobile sound, namely, quality of comfort, powerfulness, and acceleration. Then, the EEG features of the subjects are classified through a machine learning algorithm, by which the recognition of diversified automobile sound is realized. In addition, the Kalman smoothing and minimal redundancy maximal relevance (mRMR) algorithm is used to improve the recognition accuracy. The results show that there are differences in the neural characteristics of diversified automobile sound quality, with a positive correlation between EEG energy and sound intensity. Furthermore, by using the Kalman smoothing and mRMR algorithm, recognition accuracy is improved, and the amount of calculation is reduced. The novel idea and method to explore the cognitive laws of automobile sound quality from the field of brain-computer interface technology are provided in this study.

Keywords: automobile sound quality, EEG, brain cognition laws, Kalman smoothing, mRMR

OPEN ACCESS

Edited by:
Jose Luis Contreras-Vidal, University of Houston, United States

Reviewed by:
Bradley Jay Edelman, Max Planck Institute of Neurobiology (MPIN), Germany
Seong-Eun Kim, Seoul National University of Science and Technology, South Korea

***Correspondence:**
Zhien Liu
lzen@whut.edu.cn

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

Received: *03 February 2021*
Accepted: *04 June 2021*
Published: *08 October 2021*

Citation:
Xie L, Lu C, Liu Z, Yan L and Xu T (2021) Study of Auditory Brain Cognition Laws-Based Recognition Method of Automobile Sound Quality. Front. Hum. Neurosci. 15:663049. doi: 10.3389/fnhum.2021.663049

INTRODUCTION

Methods that are applied to evaluate automobile sound quality mainly rely on the psychological feelings of people and cannot guarantee the universality of evaluation results (Tan and Tan, 2012). Methods of ranking, semantic differentiation (Guo et al., 2017), grade score, pairing comparison (Parizet, 2002; Ellermeier et al., 2004) are commonly used for subjective evaluation. However, when the sound qualities with similar semantics (such as "comfort," "powerfulness," and "acceleration") are designed under the dominance of sound forward design, and the traditional subjective evaluation methods are difficult to reflect the true feelings of the evaluator. In addition to inherent physical parameter characteristics of sounds, the evaluation of an evaluator for the sound is also related to their cognition, experience, and emotional state (Genuit, 2004). Therefore, it is necessary to introduce a new automobile sound quality evaluation method for evaluating the diversified automobile sound.

Related Works

In recent years, with the research on physiological signals in emotional computing, it has become possible to use physiological signals to evaluate automobile sound. EEG signals with high time and spatial resolution are widely used (Lin et al., 2010; Bhatti et al., 2016; Geethanjali et al., 2018).

The analysis of EEG signals is challenging, and the analysis of EEG signals in the field of emotion recognition relies on data pre-processing, feature extraction (Tsang et al., 2010; Kai et al., 2016; Poikonen et al., 2016), and feature classification. Feature extraction is crucial to ensure recognition performance. Only by selecting EEG features closely related to the purpose of research can effectively meet the performance of recognition (Nishimura and Mitsukura, 2013; Sheykhivand et al., 2020). Some studies indicated that rhythm characteristic of EEG can reflect human brain activities, which are δ (1–4 Hz), θ (4–8 Hz), α (8–12 Hz), β (12–30 Hz), and γ (>30 Hz) (Knyazev, 2012; Zheng and Lu, 2015). Chen et al. (2021) proposed an EEG physiological acoustic index to evaluate subjective annoyance by comparing EEG rhythm characteristics and the change in the trend of subjective annoyance index data. Li et al. (2014) used white noise and pure tone as stimulus sources to study the relationship between EEG characteristic signals and subjective annoyance, and it is found that the average power of θ waves has two peaks in each brain area during steady stimulation. Ali et al. (2013) studied EEG signals under different sound pressure levels and stimulation intervals, and the study found that the θ wave voltage increased significantly because of high sound pressure level stimulation. Di and Wu (2015) showed that the average α wave power in the left frontal lobe was significantly lower than that in the right frontal lobe under the stimulation of pleasant sounds.

In the study of automotive sound quality and EEG signals, Lee and Lee (2014) introduced a new method to study human sound perception by means of EEGs, where EEG analysis and

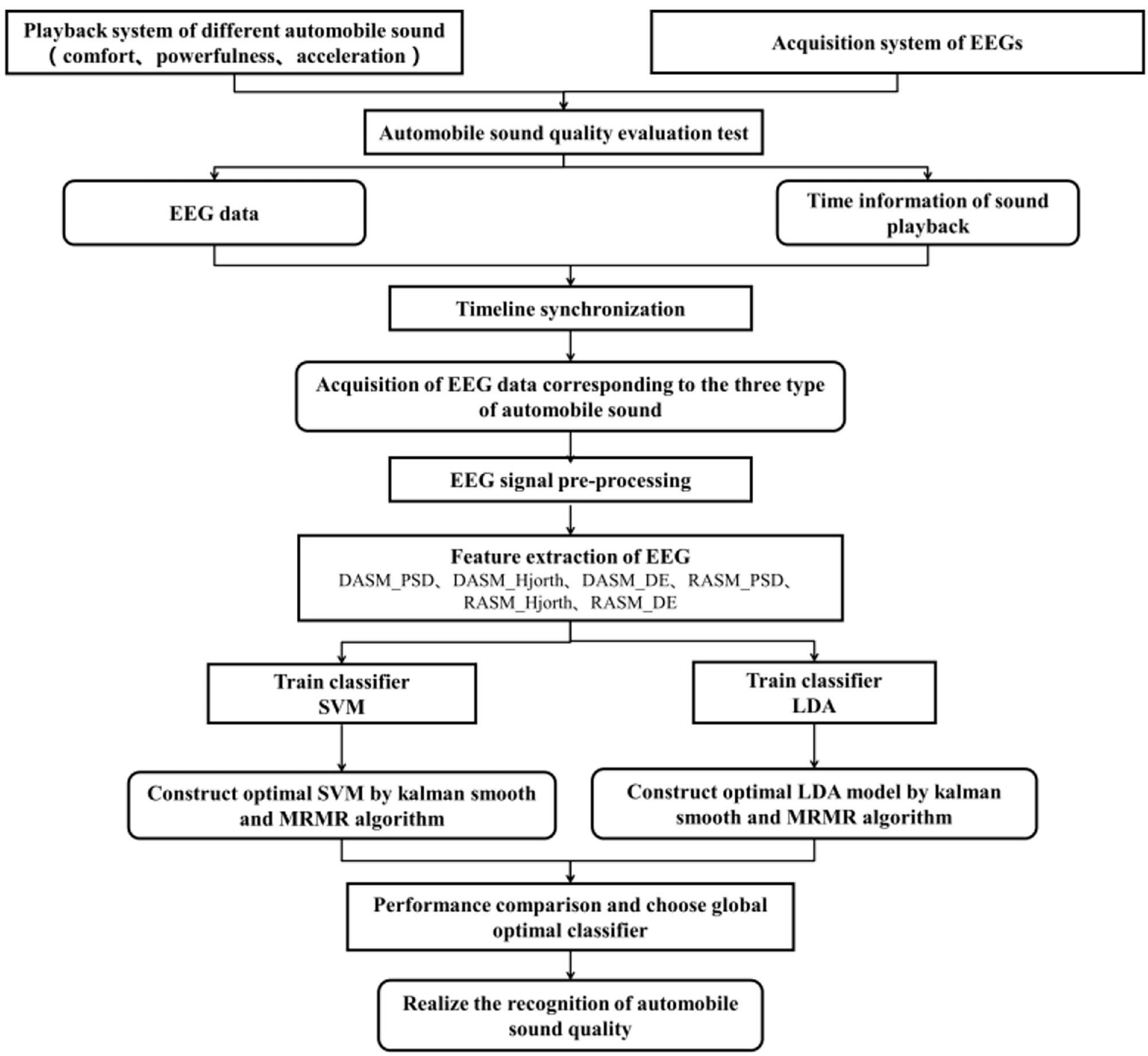


FIGURE 1 | Flow chart for the new evaluation method of car sound quality based on brain signals.

measurement were performed to demonstrate human cerebral response to car acceleration sounds and concluded that the α-wave power could serve as an objective evaluation index of automobile acceleration sounds. Lee et al. (2013) selected the α-wave to calculate the correlation between subjective evaluations of passenger car sounds and their results indicate that the intensity of the correlation between the cerebral α-wave and subjective evaluations can be determined based on the size of the correlation. Nishimura and Mitsukura (2013) put forward a group method of data handling (GMDH) to analyze the sound quality of EERs utilizing neural networks. Compared with the result efficiency of the principal component analysis (PCA), the GMDH neural network resulted in a higher recognition of the target sound quality. The above studies showed that the distinct physiological response of the human brain to sound stimuli authentically exists.

Contribution

It is difficult to distinguish automobile sounds with similar semantics by means of traditional subjective evaluations. In contrast to the application of EEG signals for emotion recognition, the study of automobile sound quality based on EEG is in infancy, the relationship between EEG feature signals and automobile sound quality is still unclear, and there is less relevant literature. However, there are related research studies on actively playing music based on EEG to improve the subjective emotions of people (Bajaj and Pachori, 2015; Kalaganis et al., 2016). Therefore, a method for mapping EEGs and diversified sound quality for decoding automobile sounds is proposed to reveal the feasibility of using EEG signals as a method of automobile sound quality evaluation, which can avoid language description. The study on decoding automobile sound types can lay the foundation of neuroscience for realizing active playback of automobile sounds based on EEGs in the future.

The auditory brain cognition laws refer to the rhythmic activities of the brain under the stimulation of the automobile sound. At present, there are no unified standards for the selection of EEG features, and it requires relevant guidance in selecting EEG features. Thus, changing the law of EEG under the stimulation of automobile sound is studied here, so as to guide the selection of EEG features. By defining three subjective evaluation indices of automobile sound quality (namely, comfort, powerfulness, and acceleration), sounds that matched with the three subjective evaluation indices are collected, The EEGs of the subjects are measured under the stimulation of three automobile sounds, respectively, in a suitable temperature and quiet environment, and the analysis of EEG data contribute to explore the cognition laws of the brain. The differential asymmetry (DASM) and rational asymmetry (RASM) features of subjects are extracted based on cognition laws, and use classification models to identify differences in automobile sound. The flow chart is shown in **Figure 1**.

Study Outline

The layout of this study is as follows: the design of the experiment is introduced in section Experiment Design. Section Methodology systematically describes the analysis methods of brain signal feature extraction, selection, and classifier. The results of data analysis are shown in section Experiment Result, including the cognitive laws of the brain under three types of automobile sounds, the use of classification models to compare the recognition accuracy differences of different features, and the optimization of model accuracy using the Kalman smoothing and mRMR algorithm. Section Discussion discusses the results of Section Experiment Result and describes the research significance of this study. Section Conclusions shows the summary and prospects of this study.

TABLE 1 | Details of the sound clips used in the EEG experiment.

No.	Labels	Sound sample sources	#Samples
1	Comfort	obtain the acceleration sound in the car under the WOT of Audi Q5, Audi A8, and FAW Toyota Prada by test	3
2	Powerfulness	obtain the acceleration sound in the car under the WOT of Lexus nx, Alfa Romeo by test; Gets the acceleration audio of Maserati president's car by video website or car game software	3
3	Acceleration	Get comfort car acceleration game simulation audio by video website and car game software	3

TABLE 2 | Characteristic distribution of evaluators.

Category	Constituent	Quantity	Percentage
Gender	Male	27	70%
	Female	12	30%
Occupation	Teacher	5	13%
	Automotive engineer	24	61%
	Postgraduate	10	26%
Age	20–29 years	26	67%
	30–39 years	5	13%
	40 years or more	8	20%
Driving experience	Yes	30	77%
	No	9	23%

EXPERIMENT DESIGN

The three types of automobile acceleration sounds are selected (namely, comfort, powerfulness, and acceleration) as inducing materials for EEG tests. These sounds that cause strong subjective and physiological changes in the subjects are mainly obtained by means of vehicle measurements, online research (such as collect acceleration sound samples of high-end automobile on website sites or from car game software), etc. **Table 1** lists the three types of automobile sounds used in the experiment. It is of significance to emphasize that these automobile sounds are divided into three parts, namely, comfort, powerfulness, and acceleration, by 39 engineers with experience in sound quality analysis, and the characteristic distribution of the 39 evaluators is shown in **Table 2**. The aim of this study is to identify three types of automobile sounds based on EEG signals. Assuming that comfort is −1, powerfulness is 0, and acceleration is 1 here, these data labels make sense when training a classifier.

Based on the experimental design and selection of subjects by Zheng and Lu (2015), a total of 15 healthy subjects are recruited, who are different from the 39 engineers. All the subjects included 11 males and four females (aged: 22.4 ± 2.53 years) who are professors or graduate students from the Wuhan University of Technology. They all have experience in automobile sound quality evaluation and ensure their optimal mental health.

Before the experiment started, the test operation procedures and specifications were relayed to all the subjects in advance, and they were instructed to properly wear high-fidelity headphones and press buttons combined with the interface prompts. Making sure that the subjects concentrate on listening to sounds and avoid obvious limb movements during the experiment is of great importance. A 64-channel AgCl electrode cap is used to collect EEG at a sampling rate of 1,000 Hz. The EEG lead distribution and electrode cap test are shown in **Figure 2**.

The three automobile sounds in each type are played randomly, and each sound is played 27 times repeatedly. There is a 5 s start prompt before each sound is played, and 10 s rest feedback after playing. A questionnaire format that the computer interface will pop up the type selection item during the 10 s rest feedback period is used, and the subjects judge which type the sound belongs to (namely, comfort, powerfulness, or acceleration). The playback process is shown in **Figure 3**.

METHODOLOGY

Feature Extraction

Combining the effective features in the field of emotion recognition, the power spectral density (PSD) (Thammasan et al., 2016), Hjorth (Jorth, 1970), and differential entropy (DE) (García-Martínez et al., 2016) are extracted as the basic EEG features in this study.

The Welch algorithm is used to set a 1-s long rectangular window with an overlap rate of 50% and obtain the PSD corresponding to different frequency bands. The Hjorth parameters, such as activity, mobility, and complexity (Vidaurre et al., 2009; Kaboli et al., 2015) are defined as

$$Activity = \mathrm{var}(X(t)) \tag{1}$$

$$Mobility = \sqrt{\frac{\mathrm{var}(\frac{dX(t)}{dt})}{\mathrm{var}(X(t))}} \tag{2}$$

$$Complexity = \frac{Mobility(\frac{dX(t)}{dt})}{Mobility(X(t))} \tag{3}$$

where var denotes the variance of the calculated $X(t)$ signal.

The DE that satisfies the Gaussian distribution is defined as (García-Martínez et al., 2016).

$$H(X) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{(x-\mu)^2}{2\sigma^2} \log \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{(x-\mu)^2}{2\sigma^2} dx = \frac{1}{2}\log 2\pi e\sigma^2 \tag{4}$$

where X means a continuous source, Gaussian distribution satisfies $N(\mu, \sigma^2)$, and π and e are a constant.

There are also several pieces of evidence that asymmetry features can well represent the cognitive laws of the human brain (Zheng et al., 2017). In this study, the DASM and RASM of 26 pairs of asymmetric electrodes are calculated, and there are six type features, which are expressed as

$$DASM\_PSD = PSD(X_{\mathrm{left}}) - \mathrm{PSD}(X_{right}) \tag{5}$$

$$DASM\_Hjorth = Hjorth(X_{\mathrm{left}}) - Hjorth(X_{right}) \tag{6}$$

$$DASM\_DE = DE(X_{\mathrm{left}}) - \mathrm{DE}(X_{right}) \tag{7}$$

and

$$RASM\_PSD = PSD(X_{\mathrm{left}})/\mathrm{PSD}(X_{right}) \tag{8}$$

$$RASM\_Hjorth = Hjorth(X_{\mathrm{left}})/Hjorth(X_{right}) \tag{9}$$

$$RASM\_DE = DE(X_{\mathrm{left}})/\mathrm{DE}(X_{right}) \tag{10}$$

The frequency is divided into five segments based on the EEG rhythm, as shown in **Figure 4**. The dimensions of DASM_PSD, DASM_Hjorth, DASM_DE, RASM_PSD, RASM_Hjorth, and RASM_DE are 130 (26 electrodes*5 rhythms), 390 (26 electrodes* 5* 3 rhythms), 130 (26 electrodes* 5 rhythms), 130 (26 electrodes* 5 rhythms), 390 (27 electrodes* 5 *3 rhythms), and 130 (27 electrodes*5 rhythms), respectively.

Feature Selection

Herein, the Kalman smoothing algorithm is used to filter out EEG components that are not associated with sounds. The purpose of Kalman smoothing is to calculate the smoothed value of the system state X_k at moment k after obtaining all observations up to time T (Cheng Y and, 2018), smoothing formula is expressed as

$$p(X_k|\mathrm{y}_{1:T}) = N(X_k|m_k^8, P_k^8) \tag{11}$$

where $T > k$, $\mathrm{y}_{1:T}$ denotes all observations in the 1∼T time period and $N(X|\mu, \sigma)$ denotes the random variable X satisfying a Gaussian distribution with mean μ and variance σ. T times forward recursion is completed from the initial time 1 to the time T, and then perform T times backward recursion from the time T to complete the Kalman smoothing process. The forward recursion process is Kalman filtering, and the state estimate m_T and covariance matrix P_T at the last time T obtained by the forward recursion are the initial state estimate m_T^8 and covariance matrix P_T^8 of the backward recursion process, namely, $m_T = m_T^8, P_T = P_T^8$.

In addition, the most common problem that is "curse of dimensionality" for pattern recognitions leads to the rapid

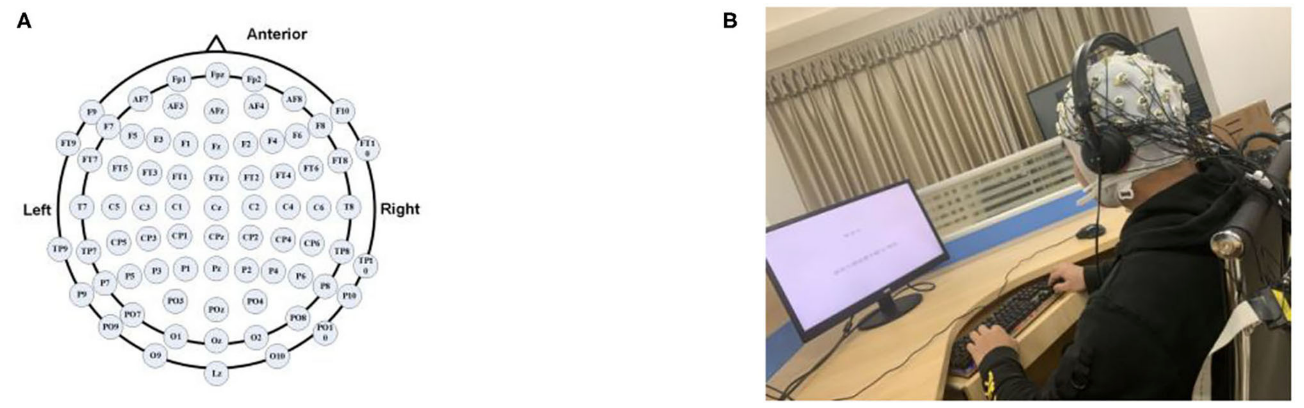


FIGURE 2 | EEG test setup: **(A)** distribution of EEG leads for 62 channels and **(B)** electrode cap test.

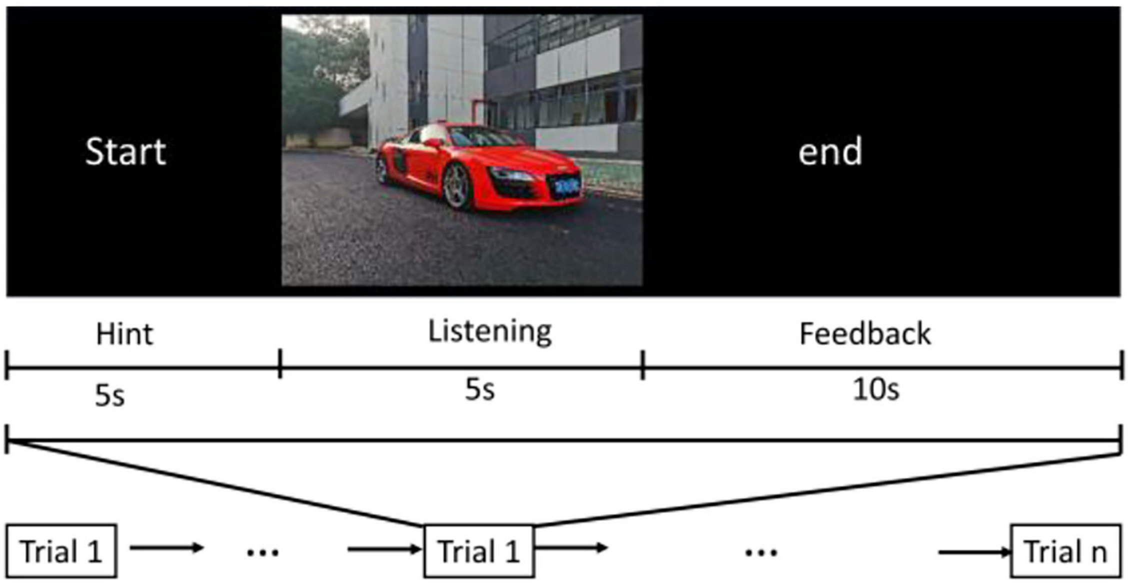


FIGURE 3 | The protocol used in the experiment for sound quality evaluation.

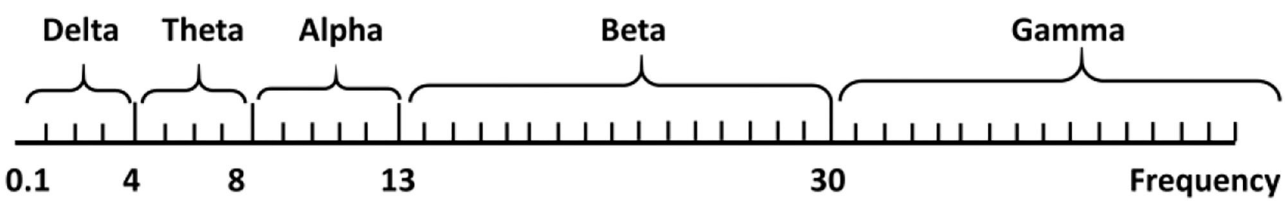


FIGURE 4 | Distribution map of EEG rhythm frequency band.

increase in computation with the increase in feature dimensions (Zheng et al., 2017). It is necessary to select EEG features after smoothing the EEG data with the target of avoiding feature redundancy, and the principal component analysis (PCA) and minimal redundancy maximal relevance (mRMR) algorithm are compared in this study.

The original domain information cannot be preserved by means of the PCA (Nakanishi et al., 2011). Hence, the mRMR algorithm is introduced to select a feature subset from EEG data here. The mRMR algorithm finds a set of features in the original feature set that is strongly correlated with the final output result (Max-Relevance), but the smallest correlation between the

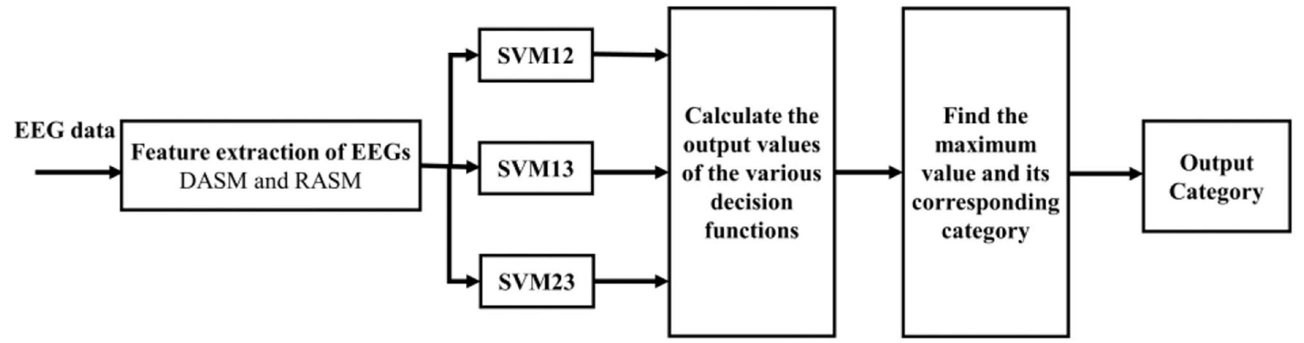


FIGURE 5 | Classification flowchart of SVM.

features (Min-Redundancy) (Peng et al., 2005). "Max-Relevance" and "Min-Redundancy" are defined as

$$\max D(S, c), D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c) \tag{12}$$

$$\min R(S), R = \frac{1}{|S|^2} \sum_{x_i, x_i \in S} I(x_i, x_j) \tag{13}$$

Combining "Max-Relevance" D with "Min-Redundancy" R, we define $\Phi(D, R)$ as

$$\max \Phi(D, R), \Phi = D - R \tag{14}$$

The approximate optimal solution can be obtained by the incremental search method, and the feature is selected by maximizing $\Phi(D, R)$.

Classifier

The reasonable design of the classifier affects the final result (Ackermann et al., 2016; Jenke et al., 2017; Hernández et al., 2018), and the linear discriminant analysis (LDA) and support vector machine (SVM) are the most common and effective classifiers. Thus, the performance differences between the LDA and SVM models are compared in this study.

The common basic idea of LDA classification assumes that every type of sample data can conform to the Gauss distribution. While a new sample arrives, it can be projected to bring their projected sample features into Gauss distribution probability density function, and then calculate its category corresponding to the peak probability.

The core idea of SVM is to find an optimal hyperplane to achieve the classification effect, and the corresponding decision function is

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{m} \alpha_i y_i K(x_i, x) + b\right) \tag{15}$$

where x_i represents the characteristics of the i-th sample, y_i represents the category of the i-th sample, and α_i the b are the calculation parameters in the SVM optimization process. The mostly used kernel function for EEG signals is the radial basis function (RBF), and the formula is as follows:

$$K(x_i, x) = e^{-\frac{\|x - x_i\|^2}{2\alpha^2}} \tag{16}$$

A "one-to-one" method was used to solve the problems of multi-classification, in which n types of training data are combined in pairs to construct n (n-1)/2 SVM. In this study, the recognition of three types of automobile sound quality is transformed into three two-classification problems. The two important SVM parameters [namely, penalty coefficient (C) and gamma] are tuned by simulation to obtain the optimal SVM model. The three sets of decision function judgment values are output, and the category with the largest judgment value is the output category of sound, namely, majority voting (Ang et al., 2012). The entire classification process is shown in **Figure 5**.

EXPERIMENT RESULT

Since the signal-to-noise ratio of EEGs is low, the original data that contain a large number of external interference noises and artifacts are necessarily preprocessed; thus, pure EEG data are extracted with the EEGLAB toolbox, mainly including EEGs (0.1–100 Hz) are captured by means of a band-pass filter, the interference band of 50 Hz is eliminated by a notch filter, the sampling rate is reset to 200 Hz, the artifacts are removed by the method of Independent Component Correlation Algorithm (ICA) and so on.

The data set input to the classification model is N*26, where 26 refers to the number of channel pairs, and N is the number of samples. There are a total of 27*9*5 = 1,215 samples (duration: 1 s) for each subject. After removal of some abnormal data, the number of EEG samples stimulated may be <1,215.

Cognitive Laws Induced by Automobile Sound

The EEG power topographic map shows the spatial distribution of power of five frequency rhythms, thereby turning complex

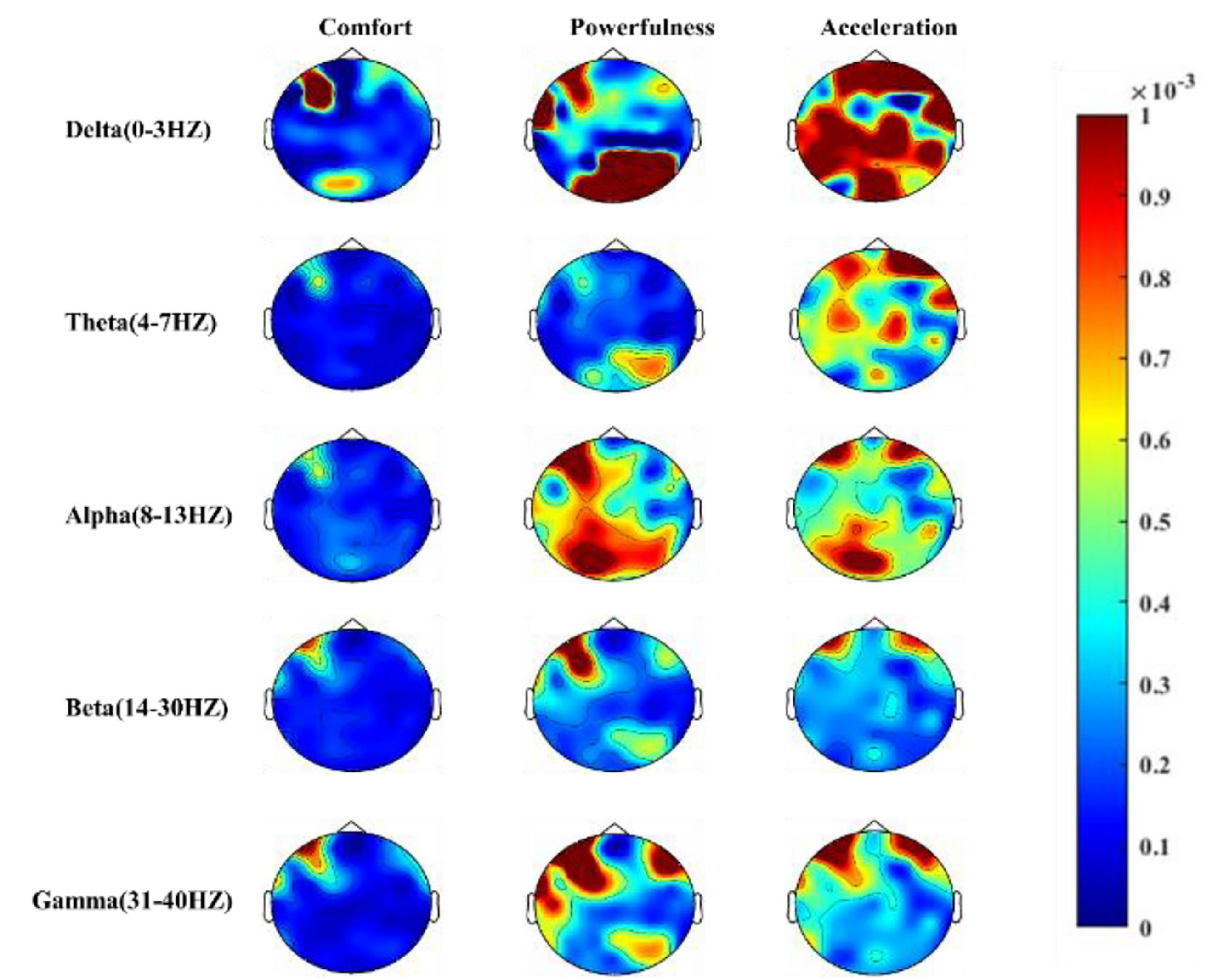


FIGURE 6 | Power topographic maps for the three types of car sounds in five rhythms.

TABLE 3 | The mean accuracy rates (%) of LDA and SVM classifiers for different features obtained from separate and total frequency bands.

Feature	Classifier	Delta	Theta	Alpha	Beta	Gamma	Total
DASM_PSD	LDA	50.63	40.03	44.34	65.17	66.17	75.01
	SVM	59.48	40.68	44.37	66.33	68.27	74.83
DASM_DE	LDA	47.56	46.78	49.33	69.71	83.74	84.83
	SVM	54.98	52.28	52.98	73.35	87.43	86.26
DASM_ Hjorth	LDA	46.83	45.30	51.19	74.05	84.24	81.47
	SVM	49.58	46.45	51.10	75.39	86.08	81.02
RASM_PSD	LDA	42.71	37.59	40.97	62.70	62.89	68.50
	SVM	49.02	39.11	41.78	63.79	64.27	69.11
RASM_DE	LDA	45.58	44.18	47.93	69.76	82.75	83.67
	SVM	51.79	48.94	51.50	73.14	**87.60**	85.49
RASM_ Hjorth	LDA	**40.10**	42.28	48.04	72.45	85.00	80.63
	SVM	44.19	43.91	49.34	74.83	86.85	81.92

Bold values = highest and lowest accuracy.

brain function changes into easy-to-follow graphs. The power topographic maps of five frequency rhythms (δ, θ, α, β, and γ) of the 15 subjects are drawled, as shown in **Figure 6**.

First, the spectrum power of the five bands under these two kinds of sound stimulation is higher than that of comfort from the perspective of a sense of powerfulness and acceleration. Based on the stimulation of powerful automobile sounds, the energy of the δ rhythm is mainly concentrated in the top and occipital areas of the bottom-right, and the energy is also more prominent in the frontal area of the upper left corner. The θ rhythm is similar to the delta rhythm but lower than δ. The energy of the α rhythm is mainly concentrated in the top area of the lower left and the frontal area of the upper left, and the β rhythm is mainly concentrated in the frontal area of the upper left, and the γ rhythm is symmetrically distributed around the frontal area.

Under the stimulation of acceleration automobile sounds, the δ rhythm energy of the entire brain is more prominent. The energy of θ and α rhythm is symmetrical in the left and right frontal regions, but the energy of θ in the central region is obvious. The energy of α in the left lower occipital region is prominent. The energy distribution of the β and γ rhythms shows a symmetrical distribution in the left and right frontal areas. As for the comfort sounds, the energy of the five frequency rhythms is obvious in the upper left frontal area.

In general, there are clear differences in the frequency band characteristics of EEG rhythm under different quality of sound stimulation.

Feature Selection

The frequency band energy of the symmetric electrode has a significant difference under the stimulations of diversified automobile sound quality; thus, the symmetrical EEG features are used as input of classifiers in this study. The LDA and SVM are used as classifiers to recognize the three types of automobile sounds, a 5-fold cross-validation scheme is adopted, and the accuracy of the classifier as an evaluation index of classifier performance.

Table 3 shows the mean accuracy of LDA and SVM for symmetrical EEG features (namely, DASM_PSD, DASM_Hjorth,

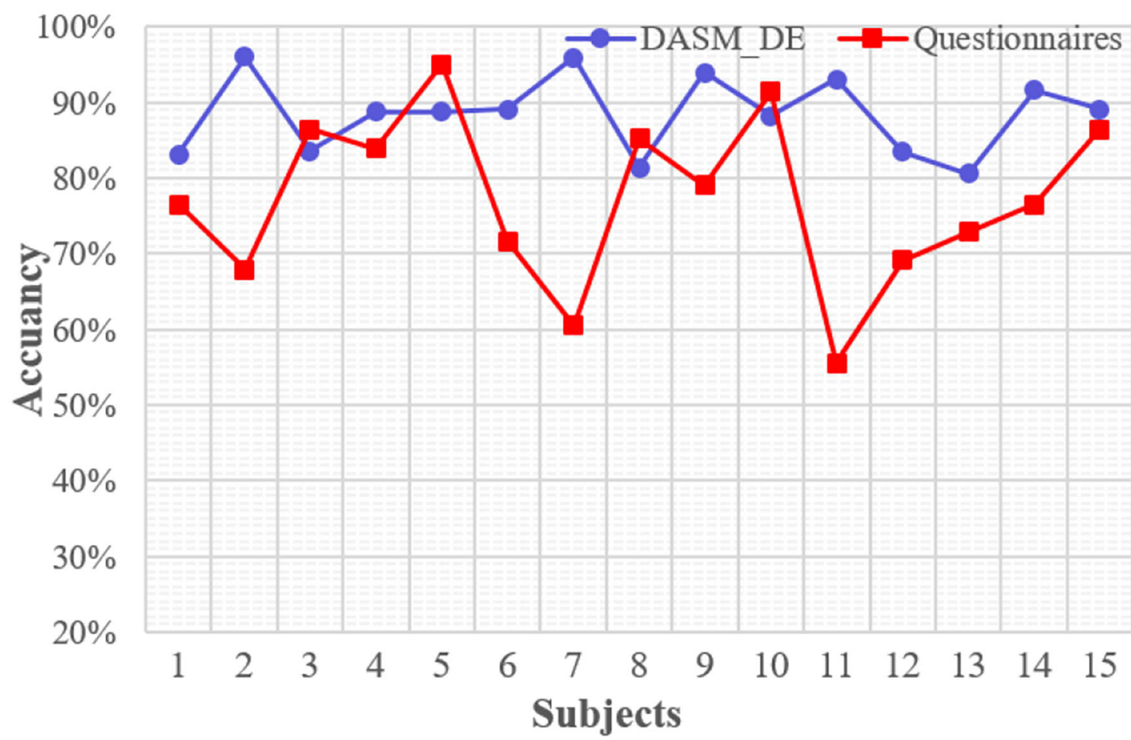


FIGURE 7 | The results of identifying the two types of automobile sounds (namely, powerfulness and acceleration) using SVM with DSAM_DE as the feature and the test subjects in form of a questionnaire.

TABLE 4 | The accuracies (%) of unsmoothing and Kalman smoothing method with RASM_PSD features of 120 dimensions as inputs and SVM as a classifier from the total frequency bands.

State	Delta (%)	Theta (%)	Alpha (%)	Beta (%)	Gamma (%)	Total (%)
Unsmoothing	49.02	39.11	41.78	63.79	64.27	69.11
Smooth	68.8	60.12	62.1	84.33	85.67	90.36
Difference	19.78	21.01	20.32	20.54	21.4	21.25

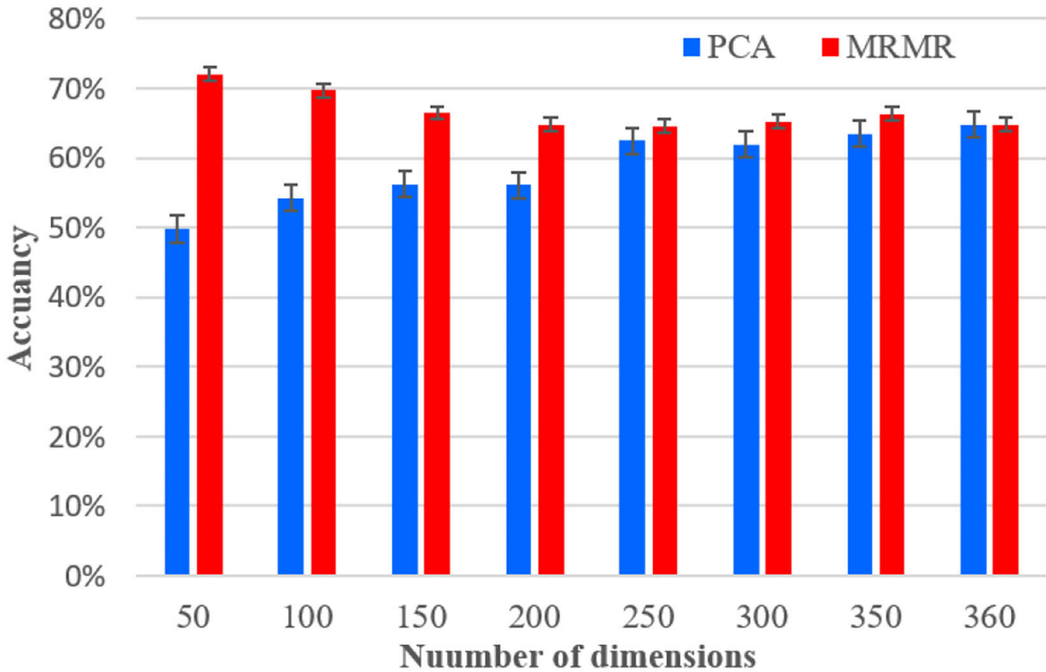


FIGURE 8 | The average accuracies of SVM using RASM_ Hjorth obtained from total frequency bands base on PCA and mRMR for subject 12.

DASM_DE, RASM_PSD, RASM_Hjorth, and RASM_DE) obtained from the five rhythms (δ, θ, α, β, and γ) and the total frequency bands. The LDA average accuracies (%) are 75.01, 84.83, 81.47, 68.50, 83.67, and 80.63 for the six features from the total frequency bands, respectively. For SVM, the average accuracies (%) are 74.83, 86.26, 81.02, 69.11, 85.49, and 81.92. In the total frequency band, the optimal and worst accuracies (%) of the LDA classifier are 86.26 and 69.11, respectively, and for the SVM classifier 84.83 and 68.50, respectively. In the total frequency band, the best and worst accuracy results appear in DASM_DE and RASM_PSD, respectively.

Further, from the classification results of the five rhythms, the LDA classifier has the lowest accuracy with 40.1% in δ rhythms with RASM_Hjorth as the feature. The accuracy up to 87.6% of the SVM classifier is the highest in the γ rhythms with RASM_DE as the feature.

The method of one-factor analysis of variance is used to study the statistical significance of the data, where the results of DE and Hjorth are better than those of PSD, and the difference in classifier performance between LDA and SVM is not apparent ($p > 0.05$). There is a significant difference in classification accuracy ($p < 0.05$) in diverse rhythms, and the accuracies of β and γ bands are significantly better than those of the three rhythms. The classification accuracy of δ, θ, and α is not totally different ($p = 0.04462$).

The powerfulness and acceleration are semantically similar. It is difficult to distinguish the difference based on subjective feelings during the experiment, which is susceptible to lead to confusion. **Figure 7** revealed that the semantic similarity recognition effect of automobile sound based on EEG signals is better than that of subjective questionnaire recognition method. **Figure 7** shows the results of identifying the two types of automobile sounds (namely, powerfulness and acceleration) using SVM with DSAM_DE as the feature and the test subjects in form of a questionnaire. It is obvious that the accuracy of the questionnaire is lower than machine learning recognition, and the average accuracy of SVM is about 11% higher than the questionnaire. It is worth explaining that the subjective recognition rate of the two other pairwise comparisons (comfort vs. powerfulness and comfort vs. acceleration) is both high, and the average accuracy rate is about 90%, which makes it difficult to reflect the advantages of EEG recognition.

Optimization of Classifier Accuracy

Firstly, the Kalman smoothing algorithm introduced in section Feature Selection is used here to remove noise that is not related to the desired signal, and the RASM_PSD features of 120 dimensions as inputs, SVM as a classifier. Second, the PCA and mRMR are compared with RASM_Hjorth features of 360 dimensions as inputs and SVM as a classifier.

Table 4 compares the accuracy of the algorithm using Kalman smoothing and without any smoothing algorithm in different rhythms. The accuracy (%) of the unsmoothing method and the Kalman smoothing method in five rhythms is 49.02/68.8, 39.11/60.12, 41.78/62.1, 63.79/84.33, 64.27/85.67, and 69.11/90.36. It is obvious that the accuracy of the Kalman smoothing algorithm method is significantly better than unsmoothing ($p < 0.05$), and the accuracy of the Kalman smoothing method is improved by 19.78% in δ rhythms and 21.4% in γ rhythms. The above results showed that feature smoothing can effectively improve the recognition accuracy.

Figure 8 compares the impact of dimension reduction using PCA and MRMR algorithms on model precision performance, in which the dimension of the model is reduced from 350 to 50 dimensions with 50 intervals. It is clear that the usage of the PCA algorithm, which can reduce the dimensionality, does not significantly improve the accuracy. The accuracy rate drops from 64.8 to 49.8% when the dimensionality reduced to 50, and it reaches 62.5% at 250 dimension, which is lower than the original 360 dimension of 1.7%. However, the mRMR algorithm can not only reduce the dimensionality, but also improve the accuracy of the classifier, the accuracy using the mRMR algorithm reached the local maximum (72.00%) at 50 dimension, which is 7.2% higher than the original 360 dimension. Moreover, the accuracy improved significantly when the dimension is 50, 100, and 150, and the dimensionality reduction is not obvious when the dimension is >150.

DISCUSSION

This study demonstrates the feasibility of EEG-based recognition of the diversified sound quality of the automobile. Several important issues are explored.

Some studies showed that the brain waves in a certain rhythm band are indeed aroused (Lee et al., 2013; Lee and Lee, 2014) under the stimulation of automobile sounds. As shown in **Figure 6**, there are frequency band differences in brain cognition under the stimulation of different sounds, which is specifically reflected in positive the correlation between EEG energy and sound energy intensity. The recognition of automobile sound quality is improved based on frequency band characteristics, which can well reflect the laws of brain cognition. Some literature has proved that the frontal area is closely related to human brain cognition (Saxe, 2006; Shamay-Tsoory and Aharon-Peretz, 2007), and there is a large proportion of energy in the frontal area under musical stimulation (Sammler et al., 2010; Di and Wu, 2015). Therefore, the results shown in **Figure 6** of this study provide further evidence that the cognition laws in the frontal portion of the human brain can indeed be aroused by automobile, so as to guide the selection of EEG features.

The DASM has better classification accuracy than RASM, which is consistent with the conclusion of the literature (Lin et al., 2010). Among the three basic features (PSD, Hjorth, DE), DE has the best classification performance, and it is most suitable for the recognition of automobile sounds. Although the classification accuracy of DASM_DE and DASM_Hjorth is close, the dimension of DASM_DE is 1/3 of DASM_Hjorth. Among the five rhythms, the classification accuracy of the β and γ rhythms is better than the other three rhythms, which proves that the correlation between different sound quality and different rhythms of brain waves is also different. The classification accuracy of the SVM model is slightly better than LDA, but SVM has the advantages of a small number of training sets, fast training speed, and high accuracy. The best accuracy of motion classification (82.29% ± 3.06%) is obtained by SVM, as demonstrated in the literature in both Lin et al. (2010) and Hadjidimitriou and Hadjileontiadis (2012), which are both similar to our study.

The comfortable sound is light and natural, and the sound pressure level is small. On the contrary, the other pairs are powerful, booming, and exciting, and the topographic map corresponding to the comfort as shown in **Figure 6** differs significantly from the other two types. For experienced automotive engineers, it is easy to distinguish the sound characteristic difference between comfort and powerfulness (or acceleration), but it is difficult to distinguish the difference between the powerfulness and acceleration sounds. In **Figure 7**, compared with recognizing sounds based on subjective feelings, using the classification model has higher recognition accuracy based on EEG characteristics. The literature (Nakanishi et al., 2011) verified the difference of EEG between three kinds of acoustic quality by using PCA and FDA in a similar way to this study. In which, the result proved that they can obtain the information that they cannot obtain from questionnaires by EEG. It is possible that the change of subjective emotion is provoked by the stimulation of the automobile sounds. However, it is not yet clear which emotion it is related to and it is the next step in the research.

As discussed in section Feature Selection, the Kalman smoothing algorithm can effectively improve the recognition accuracy and confirm that feature smoothing plays an important role in EEG-based recognition. In **Figure 8**, it is obvious that the mRMR algorithm is an effective method to optimize the accuracy of recognition, which retains the original information, such as electrode channels and frequency bands, while reducing the complexity of calculations. In the literature (Zheng et al., 2017), the mRMR algorithm was also used to achieve dimensionality reduction for improving recognition accuracy of emotion, which improves the accuracy by 14.41%.

The main contributions of this study to sound quality recognition from EEG can be summarized as follows: (1) an EEG signal acquisition test paradigm is designed based on automobile sounds, which provide experimental guidance for studying the correlation between automobile sounds and EEG signals; (2) it was systematically described the processing process of EEG data from three aspects: feature extraction, feature selection, and pattern recognition and proves that the selection of EEG features, the smoothing and dimensionality reduction of data, and the reasonable design of classifier are crucial for the recognition of sounds; (3) this study confirms that the neural characteristics of the three types of automobile sounds do exist, and the SVM can effectively identify the three types of automobile sounds through the input of the DASM_DE of γ rhythm; and (4) this research takes the brain–computer interface technology as the breakthrough point and introduces the physiological features of EEG to recognize the automobile sound quality innovatively.

CONCLUSIONS

The objective of this research is to investigate the laws of brain cognition under the stimulation of diverse automobile sounds and propose an effective method to identify diversified automobile sounds. The results show that the frequency band features can well reflect the laws of brain cognition, which can effectively realize the recognition of automobile sound quality by constructing asymmetric EEG feature indices and using machine learning models. The DASM_DE of the γ rhythm is used as the input, and the accuracy of automobile sounds reached up to 86.26% by SVM. Also, it proves that the Kalman smoothing and mRMR algorithm can not only improve the recognition accuracy but also reduce the amount of model calculation. In summary, this study proposes a new method of automobile sound quality recognition from the field of brain–computer interface technology.

Future study will include further evaluation of the specific relationship between EEG signals and the inherent characteristics of automobile sounds, proposed indices that can quantify automobile sound quality, and the usage of deep learning algorithms that automatically extract the potential features of EEGs.

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 Ethical Review Committee of Wuhan University of Technology. The participants provided their written informed consent to participate in the study.

AUTHOR CONTRIBUTIONS

ZL and LX designed the data processing schema and wrote the manuscript. CL and TX designed the experiment and were involved in the data collection. LY made a great contribution to the content of the manuscript during the revision process. All authors contributed to the article and approved the submitted version.

FUNDING

This work supported by National Natural Science Foundation of China (Grant Nos. 52175111 and 61876137) and Foshan Xianhu Laboratory of the Advanced Energy Science and Technology Guangdong Laboratory (No. XHD2020-003).

REFERENCES

Ackermann, P., Kohlschein, C., Bitsch, J. A., Wehrle, K., and Jeschke, S. (2016). "EEG-based automatic emotion recognition: feature extraction, selection and classification methods," in *IEEE International Conference on E-health Networking* (IEEE).

Ali, A., Luktuke, A. S., Ramachandran, E., and Karanth, N. V. (2013). Sound quality based benchmarking methodology for vehicle interior noise. *SAE Tech. Pap*. 12. doi: 10.4271/2013-01-2853

Ang, K. K., Chin, Z. Y., Wang, C., Guan, C., and Zhang, H. (2012). Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. *Front. Neurosci*. 6:39. doi: 10.3389/fnins.2012.00039

Bajaj, V., and Pachori, R. B. (2015). *Detection of Human Emotions Using Features Based on the Multiwavelet Transform of EEG Signals*. Springer International Publishing.

Bhatti, A. M., Majid, M., Anwar, S. M., and Khan, B. (2016). Human emotion recognition and analysis in response to audio music using brain signals. *Comput. Human Behav*. 65, 267–275. doi: 10.1016/j.chb.2016.08.029

Chen, X., Lin, J., Jin, H., Huang, Y., and Liu, Z. (2021). The psychoacoustics annoyance research based on EEG rhythms for passengers in high-speed railway. *Appl. Acoustics* 171:107575. doi: 10.1016/j.apacoust.2020.107575

Cheng, Y., Wu, X., Han, H., Xie, Y., Deng, Y., and Song, T. (2018). GNSS common-view time comparison algorithm based on Kalmanfiltering and Rauch-Tung-Streibel smoothing. *J. Harbin Eng. Univ*. 39, 120–125. doi: 10.11990/jheu?201702002

Di, G. Q., and Wu, S. X. (2015). Emotion recognition from sound stimuli based on back-propagation neural networks and electroencephalograms. *J. Acoust. Soc. Am*. 138:994. doi: 10.1121/1.4927693

Ellermeier, W., Mader, M., and Daniel, P. (2004). Scaling the unpleasantness of sounds according to the BTL model: ratio-scale representation and psychoacoustical analysis. *Acta Acustica United Acustica* 90, 101–107.

García-Martínez, B., Martínez-Rodrigo, A., Zangroniz Cantabrana, R., Pastor Garcia, J. M., and Alcaraz, R. (2016). Application of ntropy-based metrics to identify emotional distress from electro-encephalographic recordings. *Entropy* 18:221. doi: 10.3390/e18060221

Geethanjali, B., Adalarasu, K., Jagannath, M., and Seshadri, N. G. (2018). Music induced brain functional connectivity using EEG sensors: a study on Indian Music. *IEEE Sens. J*. 19:1499. doi: 10.1109/JSEN.2018.2873402

Genuit, K. (2004). The sound quality of vehicle interior noise: a challenge for the NVH engineers. *Vehicle Noise Vibrat*. 1, 158–168. doi: 10.1504/IJVNV.2004.004079

Guo, D., Shi, X. H., Wei Qing, H. U., Wen li, L. I., Peng, Y. I. (2017). Sound quality evaluation of driveline system of certain automobile based on semantic differential method. *China J. Highway Transport* 30, 307–314. doi: 10.19721/j.cnki.1001-7372.2017.06.014

Hadjidimitriou, S. K., and Hadjileontiadis, L. J. (2012). Toward an EEG-Based Recognition of Music Liking Using Time-Frequency Analysis. *IEEE Trans. Biomed. Eng*. 59, 3498–3510. doi: 10.1109/TBME.2012.2217495

Hernández, D. E., Trujillo, L., Z.-Flores, E., Villanueva, O. M., and Romo-Fewell, O. (2018). "Detecting epilepsy in EEG signals using time, frequency and time-frequency domain features," in *Computer Science and Engineering Theory and Applications* (Springer), 167–182.

Jenke, R., Peer, A., and Buss, M. (2017). Feature extraction and selection for emotion recognition from EEG. *IEEE Transac. Affect. Comput*. 5, 327–339. doi: 10.1109/TAFFC.2014.2339834

Jorth, B. (1970). EEG analysis based on time domain properties. *Electroencephalogr. Clin. Neurophysiol*. 29, 306–310. doi: 10.1016/0013-4694(70)90143-4

Kaboli, M., Walker, R., and Cheng, G. (2015). "In-hand object recognition via texture properties with robotic hands, artificial skin, and novel tactile descriptors," in *IEEE-RAS International Conference on Humanoid Robots* (IEEE).

Kai, S., Fujinaga, I., and Mcadams, S. (2016). A comparison of approaches to timbre descriptors in music information retrieval and music psychology. *J. New Music Res*. 45, 1–15. doi: 10.1080/09298215.2015.1132737

Kalaganis, F., Adamos, D. A., and Laskaris, N. (2016). *A Consumer BCI for Automated Music Evaluation Within A Popular on-Demand Music Streaming Service "Taking Listener's Brainwaves to Extremes"*. Cham: Springer.

Knyazev, G. G. (2012). EEG delta oscillations as a correlate of basic homeostatic and motivational processes. *Neurosci. Biobehav. Rev*. 36, 677–695. doi: 10.1016/j.neubiorev.2011.10.002

Lee, S. M., and Lee, S. K. (2014). Objective evaluation of human perception of automotive sound based on physiological signal of human brain. *Int. J. Automot. Technol.* 15, 273–282. doi: 10.1007/s12239-014-0028-4

Lee, Y. J., Shin, T. J., and Lee, S. K. (2013). Sound quality analysis of a passenger car based on electroencephalography. *J. Mech. Sci. Technol*. 27, 319–325. doi: 10.1007/s12206-012-1248-z

Li, Z. G., Di, G. Q., and Jia, L. (2014). Relationship between electroencephalogram variation and subjective annoyance under noise exposure. *Appl. Acoustics* 75, 37–42. doi: 10.1016/j.apacoust.2013.06.011

Lin, Y. P., Wang, C. H., Jung, T. P., Wu, T. L., Jeng, S. K., Duann, J. R., et al. (2010). EEG-based emotion recognition in music listening.. *IEEE Transac. Biomed. Eng*. 57, 1798–1806. doi: 10.1109/TBME.2010.2048568

Nakanishi, M., Mitsukura, Y., and Hara, A. (2011). *EEG Analysis for Acoustic Quality Evaluation Using PCA and FDA*. Roman. IEEE.

Nishimura, K., and Mitsukura, Y. (2013). "Sound quality indicating system using EEG and GMDH-type neural network," in *2013 Asia-Pacific Signal And Information Processing Association Annual Summit And Conference (APSIPA)*.

Parizet, E. (2002). Paired comparison listening tests and circular error rates. *ACTA Acust. United Acust.* 88, 594–598.

Peng, H., Long, F., and Ding, C. (2005). *Feature Selection Based on Mutual Information: Criteria of Max-Dependency*, *Max-Relevance, and Min-Redundancy*. IEEE Computer Society.

Poikonen, H., Alluri, V., Brattico, E., Lartillot, O., Tervaniemi, M., and Huotilainen, M. (2016). Event-related brain responses while listening to entire pieces of music. *Neuroscience* 312, 58–73. doi: 10.1016/j.neuroscience.2015.10.061

Sammler, D., Grigutsch, M., Fritz, T., and Koelsch, S. (2010). Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music. *Psychophysiology* 44, 293–304. doi: 10.1111/j.1469-8986.2007.00497.x

Saxe, R. (2006). Uniquely human social cognition. *Curr. Opin. Neurobiol.* 16, 235–239. doi: 10.1016/j.conb.2006.03.001

Shamay-Tsoory, S. G., and Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: a lesion study. *Neuropsychologia* 45, 3054–3067. doi: 10.1016/j.neuropsychologia.2007.05.021

Sheykhivand, S., Mousavi, Z., Rezaii, T. Y., and Farzamnia, A. (2020). Recognizing emotions evoked by music using CNN-LSTM networks on EEG signals. *IEEE Access* 8, 139332–139345. doi: 10.1109/ACCESS.2020.3011882

Tan, Z. F., and Tan, G. P. (2012). Evaluation method of vehicle sound quality. *Appl. Mech. Mater*. 178–181, 2829–2833. doi: 10.4028/www.scientific.net/AMM.178-181.2829

Thammasan, N., Fukui, K. I., and Numao, M. (2016). "Application of deep belief networks in eeg-based dynamic music-emotion recognition," in *The International Joint Conference on Neural Networks (IJCNN 2016)* (IEEE).

Tsang, C. D., Trainor, L. J., Santesso, D. L., Tasker, S. L., and Schmidt, L. A. (2010). Frontal EEG responses as a function of affective musical features. *Ann. N. Y. Acad. Sci*. 930, 439–442. doi: 10.1111/j.1749-6632.2001.tb05764.x

Vidaurre, C., Krämer, N., Blankertz, B., and Schlögl, A. (2009). Time domain parameters as a feature for EEG-based brain-computer interfaces. *Neural Netw*. 22, 1313–1319. doi: 10.1016/j.neunet.2009.07.020

Zheng, W. L., and Lu, B. L. (2015). "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," in *IEEE Transactions on Autonomous Mental Development*.

Zheng, W. L., Zhu, J. Y., and Lu, B. L. (2017). "Identifying stable patterns over time for emotion recognition from EEG," in *IEEE Transactions on Affective Computing*.

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.

PERSPECTIVE
published: 09 December 2021
doi: 10.3389/fcomp.2021.718605

Brain-Computer Interfaces and Creative Expression: Interface Considerations for Rehabilitative and Therapeutic Interactions

Stephanie M. Scott* **and Chris Raftery**

Department of Journalism and Media Communications, Colorado State University, Fort Collins, CO, United States

By translating brain signals into new kinds of outputs, Brain-Computer Interface (BCI) systems hold tremendous potential as both transformative rehabilitation and communication tools. BCIs can be considered a unique technology, in that they are able to provide a direct link between the brain and the external environment. By affording users with opportunities for communication and self-expression, BCI systems serve as a bridge between abled-bodied and disabled users, in turn reducing existing barriers between these groups. This perspective piece explores the complex shifting relationship between neuroadaptive systems and humans by foregrounding personal experience and embodied interaction as concepts through which to evaluate digital environments cultivated through the design of BCI interfaces. To underscore the importance of fostering human-centered experiences through technologically mediated interactions, this work offers a conceptual framework through which the rehabilitative and therapeutic possibilities of BCI user-system engagement could be furthered. By inviting somatic analysis towards the design of BCI interfaces and incorporating tenets of creative arts therapies practices into hybrid navigation paradigms for self-expressive applications, this work highlights the need for examining individual technological interactions as sites with meaning-making potentiality, as well as those conceived through unique exchanges based on user-specific needs for communication. Designing BCI interfaces in ways that afford users with increased options for navigation, as well as with the ability to share subjective and collective experiences, helps to redefine existing boundaries of digital and physical user-system interactions and encourages the reimagining of these systems as novel digital health tools for recovery.

Keywords: brain-computer interface, multimodal communication, digital creative therapy, social technology, hybrid interfaces

Edited by:
Anton Nijholt,
University of Twente, Netherlands

Reviewed by:
Juliet L. King,
George Washington University, United States
Melody Moore Jackson,
Georgia Institute of Technology, United States

****Correspondence:***
Stephanie M. Scott
SMS.Scott@colostate.edu

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Computer Science

Received: *01 June 2021*
Accepted: *16 November 2021*
Published: *09 December 2021*

Citation:
Scott SM and Raftery C (2021) Brain-Computer Interfaces and Creative Expression: Interface Considerations for Rehabilitative and Therapeutic Interactions. Front. Comput. Sci. 3:718605. doi: 10.3389/fcomp.2021.718605

INTRODUCTION

As the field of Brain–Computer Interface (BCI) technology continues to progress rapidly, it is anticipated to serve a vital role in future rehabilitation interventions for individuals experiencing neurological and/or movement disorders. Although there are various considerations for improving the accuracy and reliability of BCI systems, modern advancements within the field concern the potential of developing a system that allows immediate feedback for cognitive rehabilitation, as well as a suitable hybrid BCI for enabling creativity (Tan and Nijholt, 2010; Nishimura et al., 2010; Bamdad et al., 2015; Botrel et al., 2015; Kübler and Botrel, 2019; Nijholt, 2019). Recent research suggests that hybrid BCIs may improve overall BCI performance through combing different features of brain signals, which can include the use of two BCI navigational paradigms (for example: SSVEP and imagined movement), or through integrating BCI and another system (for example, an eye-tracker) (Bamdad et al., 2015; Todd et al., 2012). Hybrid BCIs can either simultaneously process input or operate these systems sequentially and may offer an application that can improve overall performance and enhance user experience (Todd et al., 2012).

Despite the potential for BCI systems to serve as both rehabilitation and communication tools, art and creative expression are often overlooked in assistive technology (AT) development (Huggins et al., 2019). Exploring BCI and neurofeedback development through introspective approaches highlights how novel integrative applications may provide opportunities to improve well-being resources for less-explored populations of users and underscores the importance of the user (Rapoport et al., 2008; Zickler et al., 2013; Kübler et al., 2013; Kübler et al., 2015; Morone et al., 2015; Daly and Huggins, 2015). Adapting technologies to enable digital creativity by combining professional art and music therapy practices with technical components of a responsive closed biofeedback loop may contribute to meaningful and self-expressive processes. As such, it is important to consider how applying digital forms of creative participation in BCI-enabled spaces *via* the use of internal and external feedback (both visual and auditory) can shape meaning, enhance experience, and support wellbeing for users. Utilizing theoretical tenets of communication and creative arts therapies disciplines to help guide development of these systems as innovative digital health tools, we invite consideration of the communities of practice that envelop social and technological uses of these technologies. In applying hybrid forms of navigation to further increase successful user-system literacy rates as well as encourage neuroplasticity regeneration, this work suggests that integrating multifaceted design perspectives into the design of BCI interfaces will support opportunities for more inclusive and diverse interactions.

TECHNOLOGICAL CONSIDERATIONS FOR HYBRID INTERFACES

Advances in neuroadaptive technologies, specifically BCIs, have the potential to become a "major tool for people with disabilities to control locomotion and communicate with surrounding environment and, consequently, improve the quality of life for many affected persons" (Bamdad et al., 2015, p. 355). Electroencephalography, (i.e., electrical field recording at the scalp) has the most potential for clinical application, as it is relatively simple compared to other options, and is cost effective (Bamdad et al., 2015). However, several aspects of this technology will need improvements to assist efforts to uncover new patterns of brain activity underlying artistic creation. (Rapoport et al., 2008; Tan and Nijholt, 2010; Bamdad et al., 2015; Huggins et al., 2019). By affording users equal opportunities to engage in creative activities for expression, these systems can serve as a bridge between abled-bodied and disabled users and help to reduce existing barriers between these groups (Todd et al., 2012; Kübler and Botrel, 2019).

Orndorff-Plunkett et al. (2017) suggest that experimental exploration of neuronal activity can benefit social neuroscience, arguing that processes such as neuro- and bio-feedback enable individuals to sense and interact with their own brain activity, from which causal conclusions in relation to individual behaviors, thoughts, perceptions, and experiences can be drawn (Orndorff-Plunkett et al., 2017). The notion of giving free and open choice of mental commands is notable since, on average, participants performed best with mental commands within a sensory modality they found more interesting and that corresponded to previous experience (Friedrich et al., 2012; Dhindsa et al., 2017a; Dhindsa et al., 2017b). More specifically, BCIs designed for artistic or creative applications, or BCIs that allow mental commands involving abstract visual or auditory imagery, may need to consider the artistic background of the user during training (Dhindsa et al., 2017b).

When incorporated into existing art-based BCI programs, designs focused on user-centered experience such as open-ended BCIs (Dhindsa et al., 2017a; Dhindsa et al., 2017b) and hybrid BCIs (Müller-Putz et al., 2015) have demonstrated the ability to uncover new patterns of brain activity underlying artistic creation and/or creative expression, further highlighting how applying insights from multiple disciplines can help identify gaps in interface design that reduce BCI usability (Kübler et al., 2013; Kübler et al., 2015; Müller-Putz et al., 2015; Cruz-Garza et al., 2019; Scott et al., 2019). Additionally, integrating neurofeedback applications within BCI offers clinical benefits to users through novel interventions grounded in psychological and neurosciences practices. The potential for BCI devices to facilitate general populations through neuro- and bio-feedback systems known as neurotherapeutic interventions "may give individuals a more active role in their own health care, utilize a holistic approach to body, mind, and spirit, are non-invasive, and elicit the body's own healing response" while also possessing the ability to inform social neuroscience and clinical communities" (King, 2016; Orndorff-Plunkett et al., 2017, p. 14; Scott and Gehrke, 2019; Scott et al., 2019).

CREATIVE APPROACHES TO INTERFACE AFFORDANCES

Acknowledging the reciprocal relationships between art, science and technology provides an opportunity to examine existing BCI interfaces from alternative perspectives, prompting additional research on how to improve, enhance, supplement, and allow BCI technologies to provide opportunities for creative expression and therapeutic care. (Wolpaw et al., 2002; Wolpaw and Wolpaw, 2012; Brunner et al., 2015). Expanding existing creative forms of expression for BCI systems in the form of digitally adapted creative arts therapies practices may provide different user groups the ability to communicate that which might otherwise go unexplored or un-interpreted, as it offers a medium for conveying subconscious emotions through new forms of self-expression within a therapeutic relationship (Stewart, 2004; Angheluta and Lee, 2011; Ehresman, 2014; King, 2018).

Research has demonstrated that integrating artistic practices into rehabilitative treatments through which patients are able to create or co-create art themselves improves patient-physician communication, facilitates patient thinking, and improve clinical outcomes (Sonke, 2016). Additionally, creative arts therapies have been shown to positively impact various psychological and physiological outcomes through the interactive processes of action and experiential-based training (King and Pascuzzi, 2018; de Witte et al., 2021). Initial creative arts therapy practices have been redefined through various interdisciplinary and collaborative efforts to maintain a focus on enhancing and humanizing the healthcare environment (Sonke, 2016). More recently, research on clinical and evidence based creative expression in the form of creative arts therapies, which includes art and music therapies, has demonstrated the potential for better understanding of specific health conditions while offering a safe and cost-effective intervention as an adjunct to traditional medical management (King et al., 2019; Scott et al., 2019; de Witte et al., 2021; King and Parada, 2021).

Exploring the connection between the brain and expression through feedback loops that initiate interactive processes, whether generated through bio-signals using a BCI system or *via* creative processes, requires an individual to become aware of information and learn how to use it in ways that enable communication. Furthermore, co-integrating both types of feedback loops within technological system designed to serve a therapeutic purpose may reveal a third type of loop that assumes a top-down approach; one which affords individuals the ability to rehabilitate the brain through guided feedback by using artistic and musical therapy as expression planes. This has the potential to help those who struggle with traditional learning practices to better communicate their experiences and promote their own processes of healing and recovery (Kaimal, 2019; King and Kaimal, 2019).

THEORETICAL PERSPECTIVES TOWARDS HUMAN-TECHNOLOGY RELATIONSHIPS

Acknowledging the relationships and interactions that develop *via* system engagement as unique exchanges between the human and nonhuman components highlights how technological artefacts influence the processes and production of scientific practices and knowledge. The process of recognizing material, social, and natural things invite analytical strategies that can assume more collective and habitable spaces for construing knowledge, as scientific discovery is bound by the material objects and things that compromise the scientific processes of experimentation and observation (Latour, 2009). By engaging an empirically open ethnomethodology that dissolves the boundaries between things that are considered “social” from those that are deemed “natural,” Latour (2009) suggests this process of relocation and re-embodiment of science through organized “networks of actants” and allows access towards conceptualizing how the environment is assembled. In-depth analyses of user-system interactions from these combined perspectives presents the added opportunity of engaging ethnographic methodologies, as well as interpretive and semiotic frameworks towards understanding user-system relationships as meaningful structures that are created and understood within cultural contexts (Geertz, 1973).

Engaging a postphenomenological approach towards articulating modes of knowledge as “embodied and situated” with mediated human-technological interactions enables analytical inquiry into the ways in which BCI engagement influences experience, shapes expression and impacts communication processes (Rosenberger and VerBeek, 2015, p. 1). Embodied relations through a mediated technology “simultaneously magnify or amplify and reduce or place aside what is experienced,” shaping user perception and translation through “non-neutral” device characteristics (Idhe, 1990, p. 76; Rosenberger and VerBeek, 2015), implying that the ways in which knowledge and experience are construed through BCI user-system engagement is of moral and ethical importance.

A deeper understanding of why this philosophical lens is important comes from recognizing how this perspective integrates tenets from two separate yet complementary paradigmatic frameworks-*critical* and *constructivist* (Rosenberger and VerBeek, 2015, p. 9). From a critical standpoint, it encourages dialogue between an investigator and subjects, and acknowledges findings as value-mediated while emphasizing that “researcher-researched relationships” should be built on mutual respect between equals. In this, BCI researchers, engineers, designers, etc., should not be objectively distanced from the users that are contributing to their research objectives, and BCIs should be developed in way that lends agency and equal voice to users. From a constructivist viewpoint, it recognizes that knowledge is built through user constructions and reconstructions, as well as through engagement and active participation to stimulate changes. This concept reinforces cooperation between a researcher and a user, weighing the power structure of what is possible through device affordances is of consequence (Lincoln and Guba, 1989; Guba and Lincoln, 1994, p. 230; Nisbet and Scheufele, 2009). As these paradigms maintain similar epistemological assumptions, this approach allows for both analytical critique and inquiry towards the multifaceted components of BCI systems, recognizing these devices as a composite product of technological affordances, mediated interactions with the technology, and the communication that occurs between the user and the system, as well as with researcher (Scott et al., 2019).

EMBODIED DISCOURSE OF SOCIALLY INTERACTIVE SPACES

Situating our conceptual sense of realities as grounded through metaphor and language illustrates the ways in which the design of BCI interfaces can guide new interpretations for user-system interactions (Lakoff and Johnson, 2003). By using the framework of distributed cognition as a lens through which to unite the concept of technologically-mediated social aspects of communication enabled by BCI systems with the broader mind-body processes, researchers can account for the ways in which understanding, knowledge, and perceptions are integrally situated within the articles, tools and people within our surrounding environments.

Acknowledging BCI interactions as embodied symbolic spaces through which both internal and external representations work together to specify the distributed representational space offers insight towards how these processes negotiate intrinsic (user) and environmental (technology) structural affordances in order to share information and build knowledge. Interactions cultivated through system engagement allows relationships to form and offers sites of meaning that can co-create knowledge potentials between "an individual mind and an external artifact and between individual minds" (Zhang and Patel, 2006, p. 333). This view invites further inclusive and perceptual consideration by offering a perspective that extends essentialist and normative assumptions towards use and development of these systems (Zhang and Norman 1994; Salomon, 1997; Sutton, 2006).

Assuming a translational method, otherwise referred to as "multimodal communicative competence" to analyze interactions between a user and a technology, allows researchers to understand the mediation effects on the communication processes (Royce, 2002, p. 192). Through considering the type of language used and the semiotic resources deployed, as well as various intrinsic and extrinsic structural affordances of the given medium, research can explore how meaning manifests within the bi-directional relationships fostered by BCI user-system interactions. Using an architectural model of physical spaces through three functions of analysis (experiential, interpersonal and textual), O'Halloran (2004) describes how a systemic-functional approach, or a "social semiotic" approach to interactions that occur between a user and a technology, can offer a way to build knowledge and enhance meaning (p. 27), and further inform "somatechnological" conceptualizations of these tools as user-specific sites for embodied communicative exchanges (Eco, 1976; Rosenberger and VerBeek, 2015, p. 21).

MEDIATION OF EXPERIENCE THROUGH TECHNOLOGY-ENABLED COMMUNICATION TOOLS

As the social means of communicating become increasingly intertwined with practices within medical industries, we find ourselves relying more on these technologies to both educate and assess our own personal health and manage our personal relationships, as a key component of technological implementation rests on the ability to effectively demonstrate how emerging innovation may function within the larger context of a technologically dependent society (Tarhini et al., 2015). Technological development is often accompanied by a process that involves modifying and altering emerging technologies in ways that improve functional considerations, measures of efficiency, ease of use, technical properties of a new device or system; however, these changes can introduce intermediary shifts that can occur surrounding existing practices and uses for a given technology. This adjustment occurs in part, because of technological characteristics that have changed, but also due to evolving user needs (Webster, 2002; Schulz, 2004; Cox and Depoe, 2015). The dynamic relationship between new and existing technologies is translational, in that the future impact of a technology is largely determined by its perceived use value as well as how it is received amongst professional and general communities.

BCIs are communication technologies as well as social technologies which require in-depth social, cultural, and technical analysis of the characteristics of the tools themselves, as well as the behavior that surrounds device use. Research analyzing BCI systems as social-communicative mediums emphasizes the potentiality of these to improve modes of agency and expression for users. However, additional consideration should be given to ethical and moral issues that can arise when technical changes are made to corresponding mediated environments. When specific technical constraints reinforced by digital architectures are placed upon the communication processes, it may lead to dependency and heteronomy amongst users (Dijck, 2013). This suggests that interface architecture plays a role in identity formation through the process of social interaction, and as such, efforts to examine the relationship between a specific environment and self-expression as it relates to BCI systems should not be conceived with a "one-size-fits-all" approach (Lowery and DeFleur, 1983; Postmes et al., 2005; Nisbet and Scheufele, 2009).

CONCLUSION

By its nature, BCI is a multidisciplinary field. Assuming an epistemological approach towards exploring the intersection of how these sophisticated technologies mediate communication, enable cognitively embodied interactions, and afford users the ability to share subjective and collective experiences can encourage novel conceptual understandings as to how new boundaries of digital and physical user-system interactions can explored and further applied (Scott et al., 2019; Hackett, 2008). With a focus on developing ways to contribute to therapeutic care, the goal of this work is to support the integration of creative art therapies into technological affordances of hybrid BCI systems, as this type of intervention could provide more engaging and expressive forms of treatment options to different populations, improve existing treatment options and access to care, and offer therapists a new treatment modality. By developing BCIs in a way that serves as a digital health intervention; one that engages brain activity and real-time interaction with therapeutic activities, it offers an application that combines creative expression with traditional neurofeedback practices to provide an alternative tool to improve emotional and physiological healing and recovery.

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.

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGMENTS

The authors would like to thank Dr. Charles Anderson and Dr. Rosa Mikeal Martey for their continued support.

REFERENCES

Angheluta, A., and Lee, B. K. (2011). Art Therapy for Chronic Pain: Applications and Future Directions. *Can. J. Couns. Psychother.* 45 (2), 112–131. Available at https://files.eric.ed.gov/fulltext/EJ930794.pdf.

Bamdad, M., Zarshenas, H., and Auais, M. A. (2015). Application of BCI Systems in Neurorehabilitation: A Scoping Review. *Disabil. Rehabil. Assistive Techn.* 10 (5), 355–364. doi:10.3109/17483107.2014.961569

Botrel, L., Holz, E. M., and Kübler, A. (2015). Brain Painting V2: Evaluation of P300-Based Brain-Computer Interface for Creative Expression by an End-User Following the User-Centered Design. *Brain Comput. Inter.* 2 (2-3), 135–149. doi:10.1080/2326263x.2015.1100038

Brunner, C., Birbaumer, N., Blankertz, B., Guger, C., Kübler, A., Mattia, D., et al. (2015). BNCI Horizon 2020: Towards a Roadmap for the Bci Community. *Brain Comput. Inter.* 2 (1), 1–10. doi:10.1080/2326263x.2015.1008956

Cox, R., and Depoe, S. (2015). "Emergence and Growth of the "Field" of Environmental Communications," in *The Routledge Handbook of Environment and Communication*. Editors A. Hansen and R. Cox (London: Routledge), 13–25.

Cruz-Garza, J. G., Chatufale, G., Robleto, D., and Contreras-Vidal, J. L. (2019). "Your Brain on Art: A New Paradigm to Study Artistic Creativity Based on the 'exquisite Corpse' Using mobile Brain-Body Imaging," in *Brain Art*. Editor A. Nijholt (Switzerland: Springer).

Daly, J. J., and Huggins, J. E. (2015). Brain-computer Interface: Current and Emerging Rehabilitation Applications. *Arch. Phys. Med. Rehabil.* 96 (3), S1–S7. doi:10.1016/j.apmr.2015.01.007

de Witte, M., Orkibi, H., Zarate, R., Karkou, V., Sajnani, N., Malhotra, B., et al. (2021). From Therapeutic Factors to Mechanisms of Change in the Creative Arts Therapies: A Scoping Review. *Front. Psychol.* 12, 2525. doi:10.3389/fpsyg.2021.678397

Dhindsa, K., Carcone, D., and Becker, S. (2017a). Toward an Open-Ended BCI: A User-Centered Coadaptive Design. *Neural Comput.* 29, 2742–2768. doi:10.1162/neco_a_01001

Dhindsa, K., Carcone, D., and Becker, S. (2017b). "A Brain-Computer Interface Based on Abstract Visual and Auditory Imagery: Evidence for an Effect of Artistic Training," in *Augmented Cognition, Enhancing Cognition and Behavior in Complex Human Environments. AC 2017, Part II, LNAI 10285*. Editors D. D. Schmorrow and C. M. Fidopiastis (Springer International Publishing AG), 313–332. doi:10.1007/978-3-319-58625-0_23

D. Tan and A. Nijholt (Editors) (2010). *Brain-computer Interfaces: Applying Our Minds to Human-Computer Interaction* (London: Springer-Verlag). doi:10.1007/978-1-84996-272-8

Eco, U. (1976). *A Theory of Semiotics*. Bloomington: Indiana University Press.

Ehresman, C. (2014). From Rendering to Remembering: Art Therapy for People with Alzheimer's Disease. *Int. J. Art Ther.* 19 (1), 43–51. doi:10.1080/17454832.2013.819023

Friedrich, E. V. C., Scherer, R., and Neuper, C. (2012). The Effect of Distinct Mental Strategies on Classification Performance for Brain-Computer Interfaces. *Int. J. Psychophysiol.* 84 (1), 86–94. doi:10.1016/j.ijpsycho.2012.01.014

Geertz, C. J. (1973). *Interpretation of Cultures*. New York: Basic Books.

Guba, E. S., and Lincoln, Y. S. (1994). "Competing Paradigms in Qualitative Research," in *Handbook of Qualitative Research*. Editors N. K. Denzin and Y. S. Lincoln (Thousand Oaks: Sage Publications), 105–117.

Hackett, E. J. (2008). *The Handbook of Science and Technology Studies*. 3rd ed. Cambridge, MA: MIT Press.

Huggins, J. E., Guger, C., Aarnoutse, E., Allison, B., Anderson, C. W., Bedrick, S., et al. (2019). Workshops of the Seventh International Brain-Computer Interface Meeting: Not Getting Lost in Translation. *Brain Comput. Inter.* 6 (3), 71–101. doi:10.1080/2326263X.2019.1697163

Idhe, D. (1990). *Technology and the Lifeworld*. Bloomington: Indiana University Press.

Kaimal, G. (2019). "Brain on Art Therapy-Understanding the Connections between Facilitated Visual Self-Expression, Health, and Well-Being," in *Mobile Brain-Body Imaging and the Neuroscience of Art, Innovation and Creativity*. Editors J. Contreras-Vidal, D. Robleto, J. Cruz-Garza, J. Azorín, and C. Nam (Cham: Springer), Vol. 10. doi:10.1007/978-3-030-24326-5_13

King, J. L., and Kaimal, G. (2019). Approaches to Research in Art Therapy Using Imaging Technologies. *Front. Hum. Neurosci.* 13, 159. doi:10.3389/fnhum.2019.00159

King, J. L., and Parada, F. J. (2021). Using mobile Brain/body Imaging to advance Research in Arts, Health, and Related Therapeutics. *Eur. J. Neurosci.*, 1–17. doi:10.1111/ejn.15313

King, J. L., and Pascuzzi, R. M. (2018). "Lateral Sclerosis and Related Disorders (ALS)," in *Art and Expressive Therapies Within a Medical Model: Clinical Applications*. Editors E.-A. Deborah and G. Morgan (New York, NY: Routledge), 121–126.

King, J. L., Kaimal, G., Konopka, L., Belkofer, C., and Strang, C. E. (2019). Practical Applications of Neuroscience-Informed Art Therapy. *Art Ther.* 36, 149–156. doi:10.1080/07421656.2019.1649549

King, J. L. (2016). *Art Therapy, Trauma, and Neuroscience: Theoretical and Practical Perspectives*. New York: Routledge.

King, J. L. (2018). Summary of Twenty-First century Great Conversations in Art, Neuroscience and Related Therapeutics. *Front. Psychol.* 9, 1428. doi:10.3389/fpsyg.2018.01428

Kübler, A., and Botrel, L. (2019). "The Making of Brain Painting—From the Idea to Daily Life Use by People in the Locked-In State," in *Brain Art*. Editor A. Nijholt (Switzerland: Springer).

Kubler, A., Holz, E., Kaufmann, T., and Zickler, C. (2013). "A User Centred Approach for Bringing BCI Controlled Applications to End-Users," in *Brain-computer Interface Systems-Recent Progress and Future Prospects*. Editor R. Fazel-Rezai (London: Tech Open Limited). doi:10.5772/55802

Kübler, A., Holz, E. M., Sellers, E. W., and Vaughan, T. M. (2015). Toward Independent home Use of Brain-Computer Interfaces: A Decision Algorithm for Selection of Potential End-Users. *Arch. Phys. Med. Rehabil.* 96 (3), S27–S32. doi:10.1016/j.apmr.2014.03.036

Lakoff, G., and Johnson, M. (2003). *Metaphors We Live by*. Chicago: University of Chicago Press.

Latour, B. (2009). Spheres and Networks: Two Ways to Reinterpret Globalization. *Harv. Des. Mag.* 30, 138–144. Available at https://hal-sciencespo.archives-ouvertes.fr/hal-01022658.

Lincoln, Y. S., and Guba, E. G. (1989). Ethics: The Failure of Positivist Science. *Rev. Higher Educ.* 12 (3), 221–240. doi:10.1353/rhe.1989.0017

Lowery, S. A., and De Fleur, M. L. (1983). "Developing Frameworks for Studying Mass Communication," in *Milestones in Mass Communication Research* (New York: Longman), 1–29.

Morone, G., Pisotta, I., Pichiorri, F., Kleih, S., Paolucci, S., Molinari, M., et al. (2015). Proof of Principle of a Brain-Computer Interface Approach to Support Poststroke Arm Rehabilitation in Hospitalized Patients: Design, Acceptability, and Usability. *Arch. Phys. Med. Rehabil.* 96 (3), S71–S78. doi:10.1016/j.apmr.2014.05.026

Muller-Putz, G., Leeb, R., Tangermann, M., Hohne, J., Kubler, A., Cincotti, F., et al. (2015). Towards Noninvasive Hybrid Brain-Computer Interfaces: Framework, Practice, Clinical Application, and beyond. *Proc. IEEE* 103 (6), 926–943. doi:10.1109/JPROC.2015.2411333

Nijholt, A. (2019). "Introduction: Brain-Computer Interfaces for Artistic Expression," in *Brain Art*. Editor A. Nijholt (Switzerland: Springer). doi:10.1007/978-3-030-14323-7_1

Nisbet, M. C., and Scheufele, D. A. (2009). What's Next for Science Communication? Promising Directions and Lingering Distractions. *Am. J. Bot.* 96, 1767–1778. doi:10.3732/ajb.0900041

Nishimura, E. M., Rapoport, E. D., Wubbels, P. M., Downs, T. H., and Downs, J. H. (2010). "Functional Near-Infrared Sensing (fNIR) and Environmental Control Applications," in *Brain-Computer Interfaces. Human-Computer Interaction Series*. Editors D. Tan and A. Nijholt (London: Springer). doi:10.1007/978-1-84996-272-8_8

O'Halloran, K. L. (2004). *Multimodal Discourse Analysis: Systemic-Functional Perspectives*. London: Continuum.

Orndorff-Plunkett, F., Singh, F., Aragón, O., and Pineda, J. (2017). Assessing the Effectiveness of Neurofeedback Training in the Context of Clinical and Social Neuroscience. *Brain Sci.* 7 (8), 95. doi:10.3390/brainsci7080095

Postmes, T., Haslam, S. A., and Swaab, R. I. (2005). Social Influence in Small Groups: An Interactive Model of Social Identity Formation. *Eur. Rev. Soc. Psychol.* 16 (1), 1–42. doi:10.1080/10463280440000062

Rapoport, E. D., Nishimura, E. M., Zadra, J. R., Wubbels, P. M., Proffitt, D. R., Downs, T. H., et al. (2008). Engaging, Non-Invasive Brain-Computer Interfaces (BCIs) for Improving Training Effectiveness & Enabling Creative Expression. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 52 (7), 591–594. doi:10.1177/154193120805200702

Rosenberger, R., and VerBeek, P. P. (2015). *Postphenomenological Investigations: Essays on Human-Technology Relations*. Lanham, MD: Lexington Books.

Royce, T. (2002). Multimodality in the TESOL Classroom: Exploring Visual-Verbal Synergy. *TESOL Q.* 36 (2), 191–205. doi:10.2307/3588330

Salomon, G. (1997). *Distributed Cognitions: Psychological and Educational Considerations*. Cambridge, UK: Cambridge University Press.

Schulz, W. (2004). Reconstructing Mediatization as an Analytical Concept. *Eur. J. Commun.* 19 (1), 87–101. doi:10.1177/0267323104040696

Scott, S. M., and Gehrke, L. (2019). "Neurofeedback during Creative Expression as a Therapeutic Tool," in *Mobile Brain-Body Imaging and the Neuroscience of Art, Innovation and Creativity*. Editors J. L. Contreras-Vidal, D. Robleto, J. G. Cruz-Garza, J. M. Azorín, and C. Nam (Switzerland: Springer International Publishing), 161–166. doi:10.1007/978-3-030-24326-5_17

Scott, S. M., Raftery, C., and Anderson, C. (2019). "Brain Art," in *Brain Art: Brain-Computer Interfaces for Artistic Expression*. Editor A. Nijholt. 1st ed. (Switzerland: Springer International Publishing). doi:10.1007/978-3-030-14323-7

Sonke, J. (2016). "Professionalizing the Arts in Healthcare Field," in *Managing Arts Programs in Healthcare*. Editor P. D. Lambert (New York: Routledge), 32–44.

Stewart, E. G. (2004). Art Therapy and Neuroscience Blend: Working with Patients Who Have Dementia. *Art Ther.* 21, 148–155. doi:10.1080/07421656.2004.10129499

Sutton, J. (2006). Distributed Cognition. *PC* 14 (2), 235–247. doi:10.1075/pc.14.2.05sut

Tarhini, A., Hone, K., and Liu, X. (2015). A Cross-Cultural Examination of the Impact of Social, Organisational and Individual Factors on Educational Technology Acceptance between British and Lebanese university Students. *Br. J. Educ. Technol.* 46, 739–755. doi:10.1111/bjet.12169

Todd, D. A., Mccullagh, P. J., Mulvenna, M. D., and Lightbody, G. (2012). "Investigating the Use of Brain- Computer Interaction to Facilitate Creativity," in Proceedings of the 3rd Augmented Human International Conference, Megève, France, March 1-8, 2012. doi:10.1145/2160125.2160144

van Dijck, J. (2013). 'You Have One Identity': Performing the Self on Facebook and LinkedIn. *Media, Cult. Soc.* 35, 199–215. doi:10.1177/0163443712468605

Webster, A. (2002). Innovative Health Technologies and the Social: Redefining Health, Medicine and the Body. *Curr. Sociol.* 50 (3), 443–457. doi:10.1177/0011392102050003009

Wolpaw, J., and Wolpaw, E. W. (2012). *Brain-computer Interfaces: Principles and Practice*. New York: Oxford University Press.

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 (6), 767–791. doi:10.1016/S1388-2457(02)00057-3

Zhang, J., and Norman, D. A. (1994). Representations in Distributed Cognitive Tasks. *Cogn. Sci.* 18, 87–122. doi:10.1207/s15516709cog1801_3

Zhang, J., and Patel, V. L. (2006). Distributed Cognition, Representation, and Affordance. *PC* 14 (2), 333–341. doi:10.1075/pc.14.2.12zha

Zickler, C., Halder, S., Kleih, S. C., Herbert, C., and Kübler, A. (2013). Brain Painting: Usability Testing According to the User-Centered Design in End Users with Severe Motor Paralysis. *Artif. Intell. Med.* 59 (2), 99–110. doi:10.1016/j.artmed.2013.08.003

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

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

frontiers in Computer Science

ORIGINAL RESEARCH
published: 17 January 2022
doi: 10.3389/fcomp.2021.652550

Detecting Fear of Heights Response to a Virtual Reality Environment Using Functional Near-Infrared Spectroscopy

Luciënne A. de With, Nattapong Thammasan and Mannes Poel*

Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, Netherlands

To enable virtual reality exposure therapy (VRET) that treats anxiety disorders by gradually exposing the patient to fear using virtual reality (VR), it is important to monitor the patient's fear levels during the exposure. Despite the evidence of a fear circuit in the brain as reflected by functional near-infrared spectroscopy (fNIRS), the measurement of fear response in highly immersive VR using fNIRS is limited, especially in combination with a head-mounted display (HMD). In particular, it is unclear to what extent fNIRS can differentiate users with and without anxiety disorders and detect fear response in a highly ecological setting using an HMD. In this study, we investigated fNIRS signals captured from participants with and without a fear of height response. To examine the extent to which fNIRS signals of both groups differ, we conducted an experiment during which participants with moderate fear of heights and participants without it were exposed to VR scenarios involving heights and no heights. The between-group statistical analysis shows that the fNIRS data of the control group and the experimental group are significantly different only in the channel located close to right frontotemporal lobe, where the grand average oxygenated hemoglobin Δ[HbO] contrast signal of the experimental group exceeds that of the control group. The within-group statistical analysis shows significant differences between the grand average Δ[HbO] contrast values during fear responses and those during no-fear responses, where the Δ[HbO] contrast values of the fear responses were significantly higher than those of the no-fear responses in the channels located towards the frontal part of the prefrontal cortex. Also, the channel located close to frontocentral lobe was found to show significant difference for the grand average deoxygenated hemoglobin contrast signals. Support vector machine-based classifier could detect fear responses at an accuracy up to 70% and 74% in subject-dependent and subject-independent classifications, respectively. The results demonstrate that cortical hemodynamic responses of a control group and an experimental group are different to a considerable extent, exhibiting the feasibility and ecological validity of the combination of VR-HMD and fNIRS to elicit and detect fear responses. This research thus paves a way toward the a brain-computer interface to effectively manipulate and control VRET.

Keywords: virtual reality exposure therapy, fNIRS, head-mounted display, fear of heights, classification, brain–computer interface

Edited by:
Aleksander Väljamäe, Tallinn University, Estonia

Reviewed by:
Felix Putze, University of Bremen, Germany
Alcyr Alves De Oliveira, Federal University of Health Sciences of Porto Alegre, Brazil

****Correspondence:***
Nattapong Thammasan
n.thammasan@utwente.nl

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Computer Science

Received: *12 January 2021*
Accepted: *25 November 2021*
Published: *17 January 2022*

Citation:
de With LA, Thammasan N and Poel M (2022) Detecting Fear of Heights Response to a Virtual Reality Environment Using Functional Near-Infrared Spectroscopy. Front. Comput. Sci. 3:652550. doi: 10.3389/fcomp.2021.652550

1 INTRODUCTION

Exposure therapy is a form of therapy that treats anxiety disorders by gradually and repeatedly exposing the client to his/her fear (Brinkman et al., 2009) in the absence of harm. This can activate the fear extinction process and was proven to be an effective intervention (Hofmann, 2008). Recently, virtual reality (VR) has been introduced to exposure therapy by the evidence that realistic virtual circumstances can have a significant influence on a person's mental state (Riva et al., 2007; Martens et al., 2019), which can pave a way to a successful exposure therapy. Among a vast variety of VR hardware, head-mounted display (HMD) has been shown to be effective in improving the sense of *presence* in a virtual environment (VE), which is the key element of effective application of VR in the mental health domain (Jerdan et al., 2018). Realistic immersive VEs enable researchers to ecologically perform experiments and invent therapy methods, leading to effective and highly ecologically valid virtual reality exposure therapy (VRET) systems (Martens et al., 2019). HMD-based VR enables an immersive VRET that makes the exposure therapy more controlled, safer, and in some cases also less expensive than traditional exposure therapy (Teo et al., 2016; Boeldt et al., 2019; Bălan et al., 2020). Furthermore, the exposure protocol can be completely standardized when using VRET, which increases the therapist's control over the stimuli and the duration of the exposure, as opposed to traditional *in vivo* exposure (Rizzo et al., 2013). Despite the higher level of control that VRET offers to the therapist, it is still a common practice that the therapist monitors the fear responses of the client (Brinkman et al., 2009). One important reason to do this is to ensure that the gradual exposure to the fear-eliciting stimuli do not overwhelm the client. Excessive exposure to situations that induce fear can, for example, cause panic attacks for the client and might therefore worsen the anxiety instead of treating it (Boeldt et al., 2019).

However, monitoring a person's fear responses while using VR has been a big challenge. The traditional option of tracking facial expressions becomes difficult when the user is wearing a VR-HMD. Subjective ratings suffer from the difficulty in verbalizing current mental state indication (Hill and Bohil, 2016) and memory bias (Rodríguez et al., 2015). Neuroimaging techniques have been recently proposed to objectively and unobtrusively measure fear response during virtual fear exposure but are limited to the use of electroencephalogram (EEG) (Hu et al., 2018; Peterson et al., 2018; Bălan et al., 2020). However, the disadvantages of EEG include susceptibility to motion artifacts and electrical signal interference which can be anticipated when a user interacts with VR technology. On the other hand, functional near-infrared spectroscopy (fNIRS) offers a recording of cortical activity in a natural mobility setting with higher spatial resolution than EEG, less susceptibility to motion artifacts and electrical noises, portability, and lightweight characteristic. These advantages substantiate the great potential for the combination of VR-HMD and fNIRS, which has been recently demonstrated in a bisection task (Seraglia et al., 2011), the assessment of prospective memory (Dong et al., 2017; Dong et al., 2018), the processing of racial stereotypes (Kim et al., 2019), performance monitoring during training (Hudak et al., 2017), and a neurofeedback system to support attention (Aksoy et al., 2019). However, the feasibility and ecological validity of using fNIRS to measure fear response during virtual fear exposure is still unexplored.

The neural mechanisms underpinning the fear circuit have been widely researched. The majority of fNIRS studies on cortical responses to fear-invoking stimuli report an increase in cortical activations in the parietal cortex (Köchel et al., 2013; Zhang et al., 2017) or the prefrontal cortex (PFC) (Glotzbach et al., 2011; Roos et al., 2011; Ma et al., 2013; Landowska, 2018; Rosenbaum et al., 2020) during fearful stimulation. PFC areas in which significant activations were found include the left PFC (Ma et al., 2013), dorsolateral PFC (dlPFC), anterior PFC (Landowska, 2018), left dlPFC, and left ventrolateral PFC (vlPFC) (Rosenbaum et al., 2020). The studies that found activations in the parietal cortex presented subjects to fearful and neutral sounds. Decreased chromophores deoxygenated hemoglobin (HbR) concentration changes (Köchel et al., 2013) and higher oxygenated hemoglobin (HbO) concentration changes (Zhang et al., 2017) were found when subjects were listening to fearful sounds as compared to neutral sounds. The areas with significant activations include the (right) supramarginal gyrus and the right superior temporal gyrus. The studies that found an increased cortical activation in the PFC exposed their subjects to spiders (Rosenbaum et al., 2020), fearful faces (Glotzbach et al., 2011; Roos et al., 2011), or a fear-learning experiment based on shocks (Ma et al., 2013). A recent fNIRS study observed decreased HbO concentration changes in the dlPFC and anterior PFC when participants with moderate acrophobia were exposed to a cave VE that displayed artificial heights (Landowska et al., 2018). The effect was intense during the first exposure session, but the learning process on coping with fear responses affected the following sessions. In general, the majority of fNIRS studies reported increased HbO concentration changes in the PFC when subjects were exposed to the fearful stimuli as compared to the control situations and occasionally reported the complementary decrease in HbR concentration changes (Glotzbach et al., 2011). The increment of HbO concentration changes is also in line with other neuroimaging studies beyond fNIRS that found increased cortical activity in the PFC as fearful responses (Lange et al., 2003; Nomura et al., 2004) of healthy subjects, while the activity in the amygdala is inversely related (Nomura et al., 2004). In contrast, patients with anxiety disorder show decreased activity in the PFC in response to fearful stimuli and increased activity in the amygdala (Etkin and Wager, 2007; Shin and Liberzon, 2010; Price et al., 2011). It is thus evident that the PFC plays an important role in mediating fear responses (Landowska et al., 2018) and is known as a major component of the cognitive control network (Rosenbaum et al., 2020).

Despite evidence of PFC activity due to the fear circuit as reflected by fNIRS signals, little is known about fear responses in highly immersive VR, especially when using HMD. The current study investigates the possibility of inducing and detecting fear responses in VR-HMD using fNIRS. Specifically, we are interested in inducing and detecting a fear of heights response, which is one of the most prevailing types of human fear which can be reproduced in VR, alongside (Garcia-Palacios et al., 2002;

Miloff et al., 2019; Lindner et al., 2020), fear of flying (Rothbaum et al., 2000; Maltby et al., 2002; Rothbaum et al., 2006), fear of driving (Wald and Taylor, 2000), and even post-traumatic stress disorders (Rothbaum et al., 2001; Difede and Hoffman, 2002; Gerardi et al., 2008; Rothbaum et al., 2014). Brain studies on fear of heights using fNIRS have been done in VE (Emmelkamp et al., 2001; Donker et al., 2018; Freeman et al., 2018; Gromer et al., 2018), but the previous works recruited participants either with or without fear of heights. The study in VR-HMD remains unexplored and is the main objective in this study, where we aimed to recruit participants both with and without fear of heights to allow a comparison between groups. Our first research question is as follows:

1) *To what extent do the fNIRS signals captured from participants with a fear of heights response and participants without it differ?*

To answer this question, we invited both participants with fear of heights (experimental group) and participants without fear of heights (control group) to participate in our experiment, during which they were exposed to virtual height and virtual ground conditions. It was hypothesized that the virtual heights will cause a fear response for the experimental group but does not cause a fear response for the control group. Furthermore, it was hypothesized that the ground condition does not cause a fear response for any of the groups.

In addition, we aimed to train simple machine learning classifiers to automatically detect fear responses of the experimental group from fNIRS signals, which has not been done in previous works. Our second research question is as follows:

2) *To what extent can a person's fear of heights response to a virtual reality environment be detected by a simple machine learning model using fNIRS data?*

To answer this question, we trained and tested linear classifiers in subject-dependent and subject-independent ways on the data of the experimental group and evaluated the performance in distinguishing ground-condition and height-condition data. Our first attempt to achieve a successful classification of different fNIRS responses to fear of heights elicited in VR-HMD would exhibit ecological validity of combining both components, serving as a baseline toward a practical and effective VRET in the future improvement.

2 MATERIALS AND METHODS

2.1 Participants

Two different groups of participants were recruited and pre-screened by the Acrophobia Questionnaire (AQ), consisting of 20 items that are rated on a seven-point Likert scale, ranging from *not anxious at all* to *extremely anxious* (Cohen, 1977; Antony, 2001). Only participants with fear of heights who scored higher than 35 were invited to participate as the experimental group. On the other hand, only participants without fear of heights who scored lower than 20 were invited to participate as the control group (Gromer et al., 2018). Accordingly, 20 participants (nine females, age = 26.10 ± 10.47 years) in the experimental group reported a high AQ score (52.40 ± 11.47), and 21 other participants (nine females, age = 22.95 ± 2.11 years) in the control group reported a low AQ score (9.71 ± 5.89). None of the participants suffered from anxiety disorders.

2.2 Tasks and Procedure

The study was approved by the institutional ethics committee of University of Twente (reference number: RP 2020-76). All procedures were in accordance with the Helsinki Declaration. After confirming the eligibility of the participants and obtaining written informed consent, the participants were introduced to the Oculus Rift S [1], which is a VR-HMD with six degrees of freedom enabling tracking of head rotations and translations (forward/backward, left/right, up/down). Therefore, the participants were able to look around in the VEs by simply rotating their head and to walk around by moving their body in the physical world. The researcher demonstrated how the VR-HMD should be adjusted to fit the head. Hand-held controllers were given to be held during the experiment to make the tracking of the VR-HMD more reliable, but the participants were not allowed to use the controllers. To familiarize the participants with the VE, the HMD, and holding the controller, a practice round with an example VE similar to the ground condition was included and continued until the participant indicated satisfactory familiarity. Then, the VR-HMD was removed, and the participants were fitted with the fNIRS headset. The fNIRS system was calibrated, and the signal quality was assessed visually by the researcher. Afterwards, the participants were asked to stand in a designated place and to fit the VR-HMD themselves. The straps of the VR-HMD were loosened as much as possible to reduce the risk of optode displacement. **Figure 2** shows a participant wearing both the fNIRS headset and the VR-HMD. After preparation, the participants were asked to perform the task. The researcher instructed the participants about the maximum level of movement which they were allowed to perform in order to minimize the motion artifacts in the fNIRS signal. Although the VR-HMD provides the possibility to walk around in the VE, the participants were instructed to refrain. Instead of moving, they were asked to gently look around in the VE, while preventing large head movements. Additionally, they were allowed to bend forward slightly during the height condition but were asked to return to the original position after bending forward.

All participants were tested under the same procedure. There were two conditions of VEs: ground condition and height condition. Each condition was presented alternately for five trials, each of which lasted 30 s and was preceded with a baseline period of 20 s, in which neither visual nor auditory stimuli were presented and the participants were instructed to relax and avoid active thinking. In the ground condition, the participants were virtually standing on a sidewalk or square in

[1] https://www.oculus.com/rift-s/

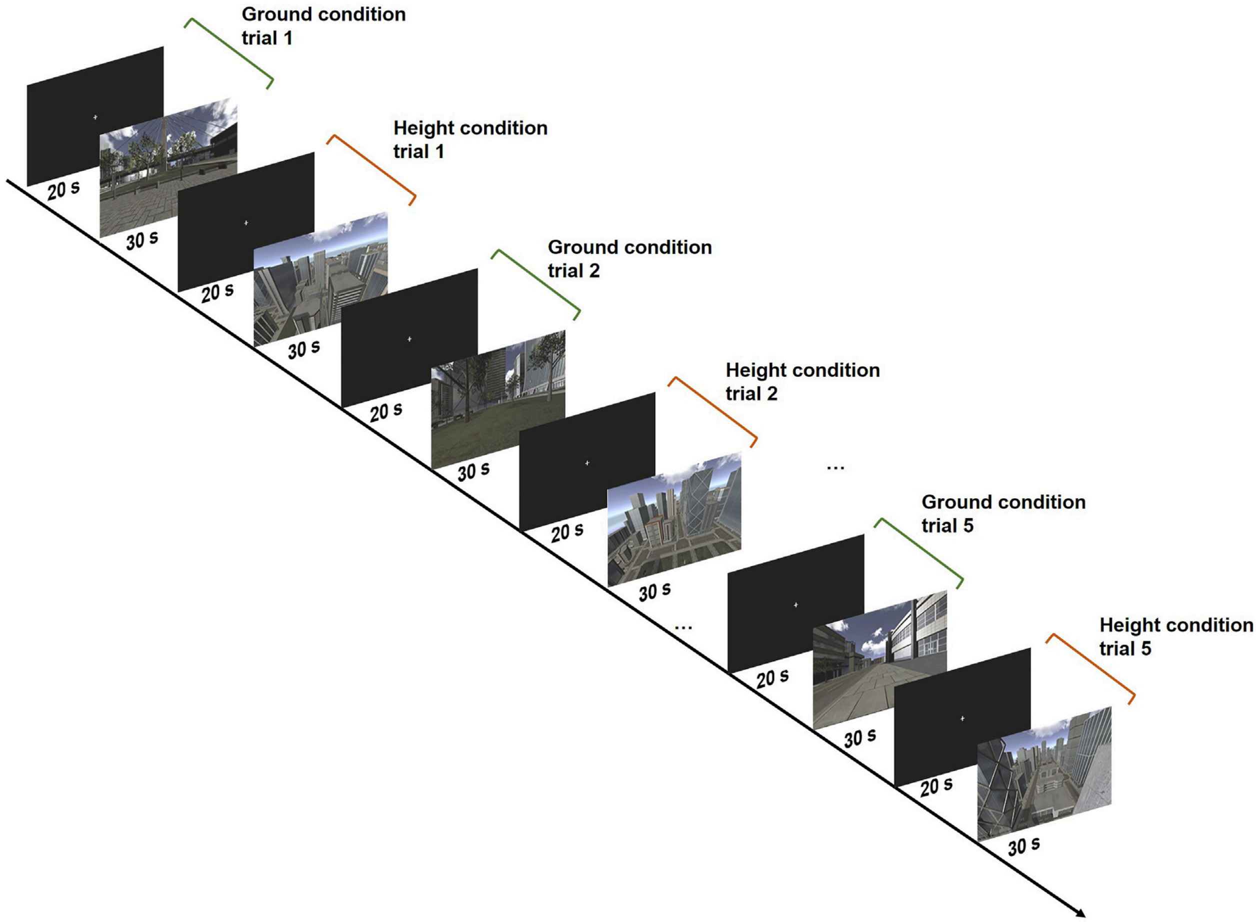


FIGURE 1 | The experimental design showing the exemplified VE scenes of the ground condition and height condition.

the middle of a city in the VE, while in the height condition, the participants were virtually standing by the rooftop of a high building. The VEs were created using the Unity development platform [2]. See **Figure 1** for examples of the scenes. Both conditions were accompanied by city sounds, to increase the immersiveness of the experience.

After the experiment, the participants were asked to rate their perceived feelings of distress or fear using the Subjective Units of Distress Scale (SUDS) (Wolpe, 1969) during ground and height conditions on an 11-point Likert scale ranging from 0 (*no distress/anxiety*) to 100 (*worst distress/anxiety that you have ever felt*). The SUDS questionnaire is often used to assess exposure settings during cognitive behavioral treatment (Benjamin et al., 2010). Additionally, the participants were asked to fill out 14 items of the IGroup Presence Questionnaire (IPQ), which measures a person's sense of presence in VR (Schubert et al., 2001), to test if the participants felt sufficiently present in the VEs for a fear response to emerge. After that, the participants were asked to fill out the AQ to confirm the group membership; median = 0.82 (Cohen, 1977) indicates adequate test–retest reliability, suggesting that pre- and post-experiment AQ scores should be similar. Finally, a structured interview by the researcher was held to ask the participants if and when they felt fearful or any other emotions during the experiment to gather extra feedback.

2.3 fNIRS Data Acquisition

Changes in HbO and HbR concentrations were measured using the Artinis Brite 24 [3]. The Brite is a wireless continuous wave fNIRS device that can measure up to 27 channels. The near-infrared light is emitted at two nominal wavelengths: 760 and 850 nm. Cortical hemodynamic responses were measured at a sampling rate of 10 Hz. The optodes were arranged to cover a large region of the PFC, including the dlPFC, anterior PFC, and part of the vlPFC. Every emitter–detector pair had a maximum distance of 3 cm between the optodes. **Figure 2** shows the positioning of the optodes and channels on the scalp, with an overview

[2]https://unity.com/

[3]https://www.artinis.com/brite

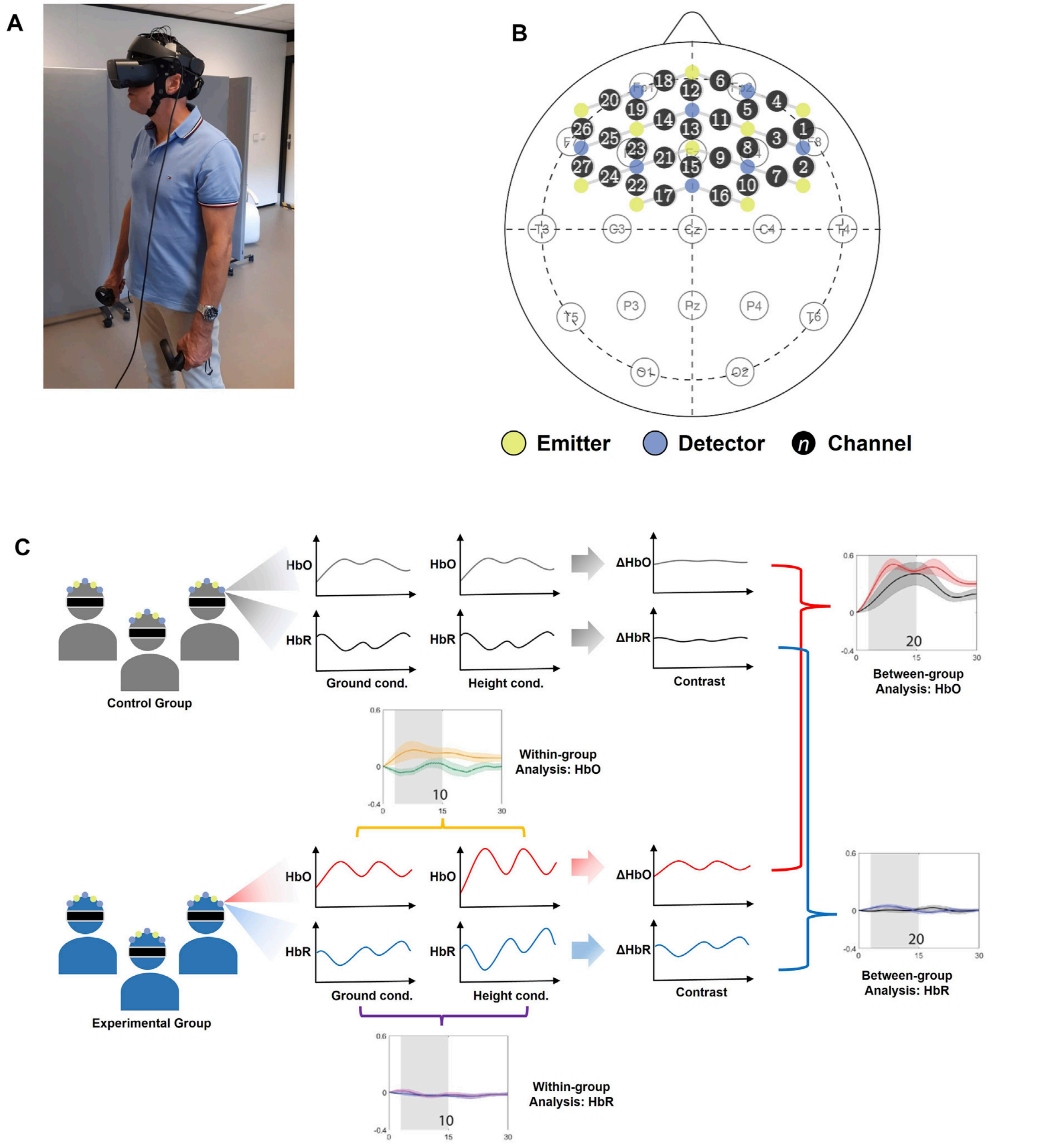


FIGURE 2 | The experimental setting; **(A)** a participant wearing the fNIRS headset and the VR-HMD during the experiment; **(B)** the positioning of the optodes projected on the layout of the 10–20 system, showing the detectors (blue), emitters (yellow), and the channels indicated by a circle with a number in it; **(C)** the overview of statistical analysis.

of the 10–20 system as a reference. In order to prevent near-infrared light being absorbed by hair, the researcher used a narrow, oblong tool to move the participant's hair to the side when it fell between an optode and the participant's scalp. Signal quality was visually validated by confirming the presence of cardiac cycles in the fNIRS signals (Hocke et al., 2018).

2.4 Data Processing

2.4.1 fNIRS Pre-processing

The fNIRS data were recorded using the Artinis Oxysoft software[4]. The raw data were converted to Δ(HbO) and

[4]https://www.artinis.com/oxysoft

Δ(HbR) signals (Chen, 2016; Pinti et al., 2019) using the modified Beer–Lambert law (Delpy et al., 1988; Scholkmann et al., 2014). After that, the data were exported to Matlab using *Oxysoft2Matlab* script and visually inspected. Channels with severe motion artifacts (usually with an amplitude of 5 μM) and channels that did not show cardiac cycles (evident by the repetitive alternation of around 0.10 μM of the amplitude) were excluded from further analysis. Motion correction was applied to the remaining channels, using the Temporal Derivative Distribution Repair procedure (Fishburn et al., 2019). After that, the correlation coefficients of every channel's Δ(HbO) and Δ(HbR) signals were calculated. Channels with a positive correlation coefficient were removed, following a previous fNIRS study suggesting that a negative correlation can be expected when the amount of motion artifacts in the signals is low (Cui et al., 2010). Then, a third-order Butterworth band-pass filter (Hocke et al., 2018; Pinti et al., 2019) with low cut-off frequency 0.01 Hz and high cut-off frequency 0.1 Hz was applied to remove physiological noise arising from breath cycles (~0.2–0.3 Hz), cardiac cycles (~1 Hz), and Mayer waves (~0.1 Hz) (Naseer and Hong, 2015; Pinti et al., 2019). The filtered signals were separated into trials and adjusted with a baseline, yielding five ground-condition trials and five height-condition trials. The main duration of trials was set from 0 to 30 s after the stimulus presentation onset, covering the entire task period of each trial. The 5-s period preceding the stimulus presentation was used as a baseline period. For each participant and each channel, the signals were grand averaged across trials in each condition.

2.5 Data Analysis

A between-group analysis was performed by investigating the significant differences of fNIRS signals between the control group and the experimental group. In order to investigate the effect of height on participants with fear of heights, a within-group analysis was performed on the data of ground-condition trials and height-condition trials of merely the experimental group to investigate significant difference of fNIRS signals in the two conditions. Statistical testing was conducted using Matlab 2020a, and the overview of the statistical analysis performed is illustrated in **Figure 2**.

2.5.1 Between-Group Analysis

To mitigate the inter-subject variability issue, we computed a contrast between the ground-condition and the height-condition grand average Δ[HbO] signals and Δ[HbR] signals for all channels and all participants. The contrast was computed by subtracting the grand average ground condition signal from the grand average height condition signal. For all of these signals, the mean over the window from 3 to 15 s post-stimulus onset was computed, following the evidences that the hemodynamic response only starts to become visible after 3 s [2.8-s lag was found (Lachert et al., 2017)] and that the hemodynamic response is most intense in the first 5–17 s after the stimulus onset (Khan et al., 2020). For every channel, a permutation test with 50,000 permutations was used to test for significant differences between the contrast signal means of the control group and the experimental group, at the significance level $\alpha = 0.05$. The permutation test was chosen as it is a non-parametric test that can be used on small sample sizes and makes no assumptions about the distribution of the data.

2.5.2 Within-Group Analyses

Similar to the between-group analysis, the grand average ground condition Δ(HbO) and Δ(HbR) signals and the grand average height condition Δ(HbO) and Δ(HbR) signals were averaged over the 3–15 s window. For every channel, a permutation test with 50,000 permutations was used to test for significant differences between the ground-condition trial means and the height-condition trial means over the 3–15 s window, at the significance level $\alpha = 0.05$.

2.5.3 Correction for Multiple Comparisons

Four statistical analyses were executed on the fNIRS data (between/within group analysis on the Δ[HbO]/Δ[HbR] data) per channel, yielding a total of $4 \times 27 = 108$ hypothesis tests from all channels. False discovery rate correction, as suggested by Genovese et al. (2002) for neuroimaging data, was executed on the 108 p-values that resulted from the statistical analyses to correct for multiple comparisons. The rate q was set to 0.05.

2.6 Classification

2.6.1 Feature Extraction

We used data from all channels that are not corrupted by movement artifacts and hardware malfunctions (as explained in **Section 2.4.1**) for classification, where the number of available channels differs across participants. Instead of extracting features per channel, we first calculated the averages of Δ[HbO] and Δ[HbR] measurements over the remaining channels, yielding the $\overline{\Delta[\mathrm{HbO}]}$ and $\overline{\Delta[\mathrm{HbR}]}$ respectively, and then extracted the features from them in the period within 3–15 s after stimulus onset. A 1-s sliding window was applied without overlap between consecutive windows in order to gain more data from the signals. Following the study of Derosière et al. (2014), the averages of $\overline{\Delta[\mathrm{HbO}]}$ and $\overline{\Delta[\mathrm{HbR}]}$ were then computed per window to represent the data of that 1-s window. Short histories of the averaged $\overline{\Delta[\mathrm{HbO}]}$ and $\overline{\Delta[\mathrm{HbR}]}$ signals of every second were also computed, such that the information arising from the changes in the signal over time could be utilized as additional features for classification, similar to Hu et al. (2012). For every observation, the current observation and the observations of the x seconds preceding the current observation were extracted (yielding a total of $x + 1$ features for each chromophore (Δ[HbO] and Δ[HbR])). In order to investigate the effect of the length of the histories, the classifiers were trained and tested on three different histories, similar to Hu et al. (2012): 1 s, 3 s, and 5 s. Note that the duration of histories to *look back* leads to different number of features, but the number of training and test instances are identical.

2.6.2 Subject-Dependent and Subject-Independent Classification

Subject-dependent classifiers were trained and tested only for the experimental group due to the clear distinction of fear responses between height-condition trials and ground-condition trials, labeled as *fear response* and *no fear response*, respectively. All

TABLE 1 | Mean scores and standard deviations of the questionnaire results for the control group and the experimental group.

Questionnaire	Control group mean (±SD)	Experimental group mean (±SD)
Pre-experiment AQ	10.80 (±5.66)	56.07 (±11.20)
Post-experiment AQ	10.73 (±6.41)	50.36 (±11.85)
SUDS ground condition	3.00 (±4.55)	6.43 (±6.02)
SUDS height condition	11.53 (±8.08)	69.86 (±11.55)
IPQ presence	4.20 (±0.86)	4.21 (±0.97)

1-s windowed data from the first six trials (consisting of three ground-condition trials and three height-condition trials) were used as training data, and the remaining windowed data from four trials were the test data. As data were extracted from 3 to 15 s after stimulus onset, this resulted in $12 \times 6 = 72$ training data instances and $12 \times 4 = 48$ test data instances from each participant. In this study, linear discriminant analysis (LDA) and support vector machines (SVM) with linear kernel, implemented in Matlab 2020a, were trained with the standard hyper-parameter settings. Specifically, a linear coefficient threshold of 0 was used with regularized LDA. Sequential minimal optimization was applied to the linear-SVM and without feature scaling. The performance in the modes of 1-, 3-, and 5-s history was measured by the accuracy. Similarly, subject-independent classifiers were trained and tested with the experimental group using leave-one-subject-out cross-validation. It can be useful in real-life VRET settings to classify unseen data from an unknown participant (Bălan et al., 2020). In order to compare the subject-dependent classification with a random classifier (50% accuracy), the 95% confidence interval is calculated for each classifier. The lower (b_l) and upper bounds (b_u) of the 95% confidence interval are based on the Wilson score interval (Wilson, 1927) and are given by the formula

$$(b_l, b_u) = \frac{1}{1+\frac{z^2}{n}}\left(\hat{p} + \frac{z^2}{2n}\right) \pm \frac{z}{1+\frac{z^2}{n}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2}{4n^2}} \quad (1)$$

where $\hat{p}$ is the estimated performance, n is the number of test samples, and z the value corresponding to the desired confidence interval. In case of the 95% confidence interval, $z = 1.96$. The advantage of the Wilson score interval is that it is asymmetric and has no overshoot or zero-width intervals unlike the traditional normal approximation.

3 RESULTS

3.1 Behavioral Results

Three participants withdrew from the experiment due to motion sickness caused by the VR-HMD. SUDS threshold at 30 was used to distinguish the feeling of *relaxation* and *fear*, where participants should report higher than this threshold when feeling fear during the experimental condition. As a result, two participants were excluded from each group due to the mismatch between the reported SUDs score and the expected range. In addition, the threshold of IPQ was set at 3 as the minimum for the feeling of presence in the VR, leaving two participants, whose scores did not surpass the threshold, out from the experimental group. Besides, the AQ scores were used to reconfirm the group membership after the experiment, resulting in one and two participants removed from the control and experimental groups, respectively. Consequently, a total of 15 participants ($n_c = 15$) remained to be part of the control group and 14 participants ($n_e = 14$) were part of the experimental group. **Table 1** shows the mean scores and standard deviations of the questionnaire results for both groups; it suggests a clear distinction between the two groups in terms of AQ scores (pre-experimental as well as post-experimental) and SUDS for the height condition. The two groups scored similarly for SUDS in ground condition and for the experienced presence in the VEs.

3.2 Statistical Analysis

3.2.1 Between-Group Analysis

Figure 3 shows the grand average Δ[HbO] traces of the contrast between the ground condition and the height condition for the two groups for every channel, with the standard error given around every trace. It is apparent that only channel 3 ($p = 0.000\,8$) generated a significant difference between the contrast Δ[HbO] means of both groups using statistical testing with the false discovery rate (FDR) correction. Meanwhile, the results from Δ[HbR] traces, as also shown in **Figure 3**, suggested that there are some channels (i.e., channels 15, 23, 24, and 27), where the grand average trace of the control group has a different pattern than that of the experimental group. However, none of them are significant after the corrected statistical testing.

3.2.2 Within-Group Analysis

Figure 4 shows the grand average Δ[HbO] traces of the ground condition and the height condition for the experimental group, with the standard error given around every trace. Apparently, the difference between the two conditions can be clearly observed, especially on the salient increase of Δ[HbO] during 3–15 s after stimulus compared to the rather constant trace in ground condition. FDR-corrected permutation test shows that the distinct patterns are significantly different in channels 1 ($p = 0.002\,2$), 2 ($p = 0.002\,2$), 3 ($p = 0.000\,01$), 4 ($p = 0.000\,3$), 6 ($p = 0.002\,2$), 11 ($p = 0.001\,6$), 12 ($p = 0.004\,1$), 14 ($p = 0.001\,6$), 18 ($p = 0.000\,9$), 20 ($p = 0.000\,02$), 23 ($p = 0.000\,02$), 25 ($p = 0.000\,1$), and 26 ($p = 0.002\,2$). In contrast, Δ[HbR] traces, as also shown in **Figure 4**, are rather flat for both conditions, whereas only the difference in channel 23 ($p = 0.001\,7$) is significant after the corrected statistical testing.

3.3 Classification

Due to motion artifacts and hardware malfunctions, some channels were excluded from the analyses in some participants; i.e., features were extracted from the remaining uncorrupted channels per participant (see **Section 2.4.1**). In this study, we trained and tested the subject-independent classifiers with the data from only of the uncorrupted Δ[HbO] channels, as many corrupted Δ[HbR] channels were excluded for

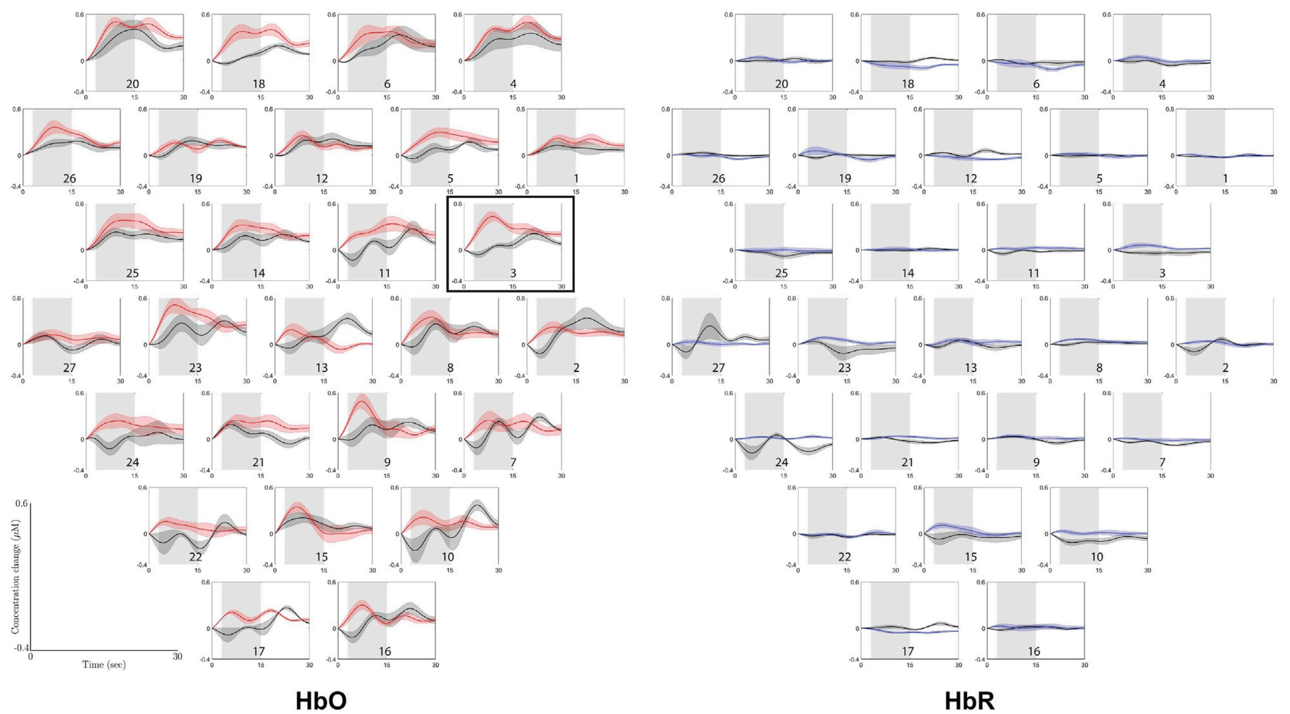


FIGURE 3 | Grand average Δ(HbO) and Δ(HbR) traces of the contrast between ground condition and height condition for the two groups: control group (black traces) and experimental group (red traces for Δ(HbO) and blue traces for Δ(HbR)), with standard deviation. The gray shaded area (3–15 s post-stimulus) is the window over which the means were taken that were used for the permutation tests. The horizontal axis represents time in seconds, ranging from 0 to 30, and the vertical axis represents concentration change in μM, ranging from −0.4 to 0.6. The plots are corresponding to channel labels, which are arranged in accordance with the optode layout that was used during the experiment, as presented in **Figure 2**. Channel numbers are labeled in every plot. The plots surrounded by the border shows the channel where a significant difference was found between the means of the control group and the experimental group. The graphs are arranged according to the optode layout that was used during the experiment, as presented in **Figure 2**.

many participants, which makes it unfeasible to train and test classifiers on these data.

3.3.1 Subject-Dependent Classification

Table 2 shows the accuracies of the subject-dependent classification and the 95% confidence interval, calculated using **Eq. 1**. The mean accuracy computed across all participants suggests that the SVM on the 3-s history performs best, with a mean accuracy of 69.69% (SD 16.94). However, the mean accuracies of the other classifiers are close to that of the 3-s history SVM, with a maximal difference of roughly 1.4%. Therefore, the amount of history taken into account in the classification seems to have a minimal effect on this metric. LDA and linear SVM achieved similar performance with a maximum of 1.2% in accuracy difference. From **Eq. 1** one can easily deduce that if the estimated performance is above 64.5% then the lower bound of the 95% confidence interval is higher than 50%. Recall that a subject-dependent classifier is tested on 48 samples; hence, $n = 48$. This implies that for most types of classifiers considered, only 7 out of the 14 subject-dependent classifiers perform significantly better than random. In order to show that the mean of the different subject-dependent classification is significantly higher than the mean of a random classifier for each participant, we take a somewhat different approach. Since the mean over all subject-dependent classification is not the outcome of a Bernoulli experiment (it is the mean over different Bernoulli experiments) we cannot apply the Wilson score interval. But the mean performance of 14 subject-dependent random classifiers is equivalent to an estimated performance of a Bernoulli experiment with 14×48 trials. Since the success rate of a random classifier is known and equal to 50%, we can estimate the 99% confidence interval ($z = 2.576$) using **Eq. 1** and is given by (45%, 55%). This means that in 99% of the cases, the observed mean of the random classification will be in this interval.

The performance also varies considerably among the different participants; while classification in participants 1, 2, and 9 achieved low accuracy, classification for participants 7 and 10 was almost perfect. In some participants, the accuracy also changed by classification methods and history by an amount of almost 15% (participant 1), while these factors had minimal impact on the accuracy in other participants (e.g., participants 7 and 10). This led us to the analysis of data distribution and its effect on the classification performance.

We investigated the data distribution in feature space of representative participants: participants 2 and 9 with relatively low classification accuracy and participant 7 with high

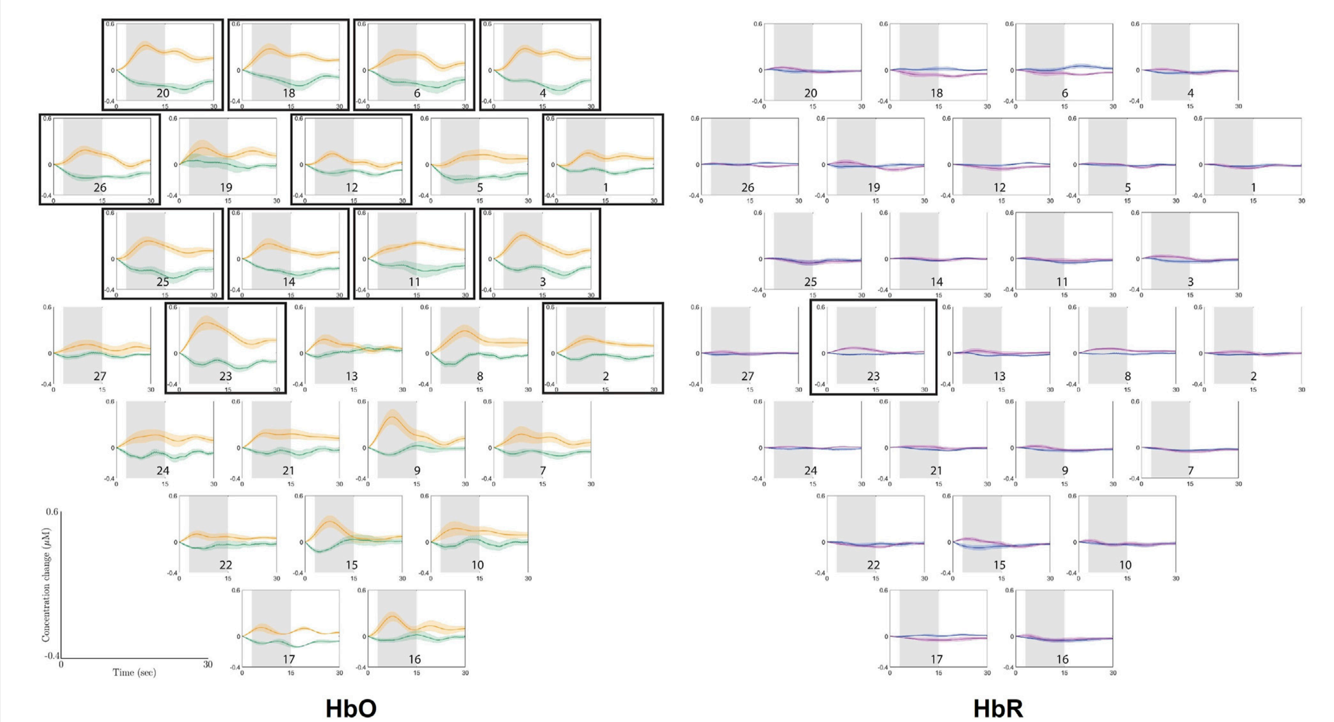


FIGURE 4 | Overview of the grand average Δ(HbO) and Δ(HbR) traces of the ground condition (green traces for Δ(HbO) and blue traces for Δ(HbR)) and height condition (orange traces for Δ(HbO) and purple traces for Δ(HbR)) of the experimental group, with standard deviation. The gray shaded area is the window over which the means were taken that were used for the permutation tests. The horizontal axis represents time in seconds, ranging from 0 to 30, and the vertical axis represents concentration change in μM, ranging from −0.4 to 0.6. The plots are corresponding to channel labels, which are arranged in accordance with the optode layout that was used during the experiment, as presented in **Figure 2**. The plots surrounded by boarders show the channels where a significant difference was found between the means of the ground condition and the height condition for the experimental group.

TABLE 2 | Accuracies and confidence intervals [lower bound, upper bound] of the subject-dependent classification.

Participant	1s history		3s history		5s history	
	LDA	SVM	LDA	SVM	LDA	SVM
1	41.67 (28.85,55.72)	56.25 (42.27,69.30)	41.67 (28.85,55.72)	56.25 (42.27,69.30)	54.17 (40.29,67.43)	50.00 (36.39,63.61)
2	52.08 (38.33,65.53)	54.17 (40.29,67.43)	43.75 (30.70,57.73)	52.08 (38.33,65.53)	43.75 (30.70,57.73)	52.08 (38.33,65.53)
3	58.33 (44.28,71.15)	64.58 (50.44,76.56)*	56.25 (42.27,69.30)	58.33 (44.28,71.15)	50.00 (36.39,63.61)	58.33 (44.28,71.15)
4	85.42 (72.84,92.75)*	85.42 (72.84,92.75)*	77.08 (63.46,86.69)*	83.33 (70.42,91.30)*	77.08 (63.46,86.69)*	81.25 (68.06,89.81)*
5	89.58 (77.83,95.47)*	83.33 (70.42,91.30)*	91.67 (80.45,96.71)*	83.33 (70.42,91.30)*	83.33 (70.42,91.30)*	81.25 (68.06,89.81)*
6	58.33 (44.28,71.15)	58.33 (44.28,71.15)	68.75 (54.67,80.05)*	64.58 (50.44,76.56)*	68.75 (54.67,80.05)*	68.75 (54.67,80.05)*
7	97.92 (89.11,99.63)*	100.00 (92.59,100.00)*	95.83 (86.02,98.85)*	100.00 (92.59,100.00)*	97.92 (89.11,99.63)*	100.00 (92.59,100.00)*
8	81.25 (68.06,89.81)*	81.25 (68.06,89.81)*	87.50 (75.30,94.14)*	87.50 (75.30,94.14)*	93.75 (83.16,97.85)*	93.75 (83.16,97.85)*
9	45.83 (32.57,59.71)	45.83 (32.57,59.71)	41.67 (28.85,55.72)	43.75 (30.70,57.73)	37.50 (25.22,51.64)	41.67 (28.85,55.72)
10	85.42 (72.84,92.75)*	83.33 (70.42,91.30)*	85.42 (72.84,92.75)*	89.58 (77.83,95.47)*	91.67 (80.45,96.71)*	87.50 (75.30,94.14)*
11	64.58 (50.44,76.56)*	60.42 (46.31,72.98)	64.58 (50.44,76.56)*	54.17 (40.29,67.43)	64.58 (50.44,76.56)*	50.00 (36.39,63.61)
12	75.00 (61.22,85.08)*	77.78 (64.22,87.22)*	83.33 (70.42,91.30)*	77.78 (64.22,87.22)*	83.33 (70.42,91.30)*	77.78 (64.22,87.22)*
13	62.50 (48.36,74.78)	60.42 (46.31,72.98)	62.50 (48.36,74.78)	62.50 (48.36,74.78)	56.25 (42.27,69.30)	62.50 (48.36,74.78)
14	62.50 (48.36,74.78)	58.33 (44.28,71.15)	58.33 (44.28,71.15)	62.50 (48.36,74.78)	54.17 (40.29,67.43)	56.25 (42.27,69.30)
Mean (±SD)	68.60 (64.99,72.00)* (±17.29)	69.25 (65.66,72.62)* (±15.64)	68.45 (64.84,71.85)* (±18.77)	69.69 (66.11,73.04)* (±16.94)	68.30 (64.69,71.71)* (±19.70)	68.65 (65.04,72.04)* (±18.32)

** indicates the performance for which the lower bound of its confidence interval is larger than 50%.*

classification performance. Specifically, principal component analysis (PCA) was applied to 1-s history data, and we visualized the distribution of data projected on the first and second principal components (PCs) to investigate if the training and test data are identically distributed as shown in **Figure 5**. Training and test data are represented in different

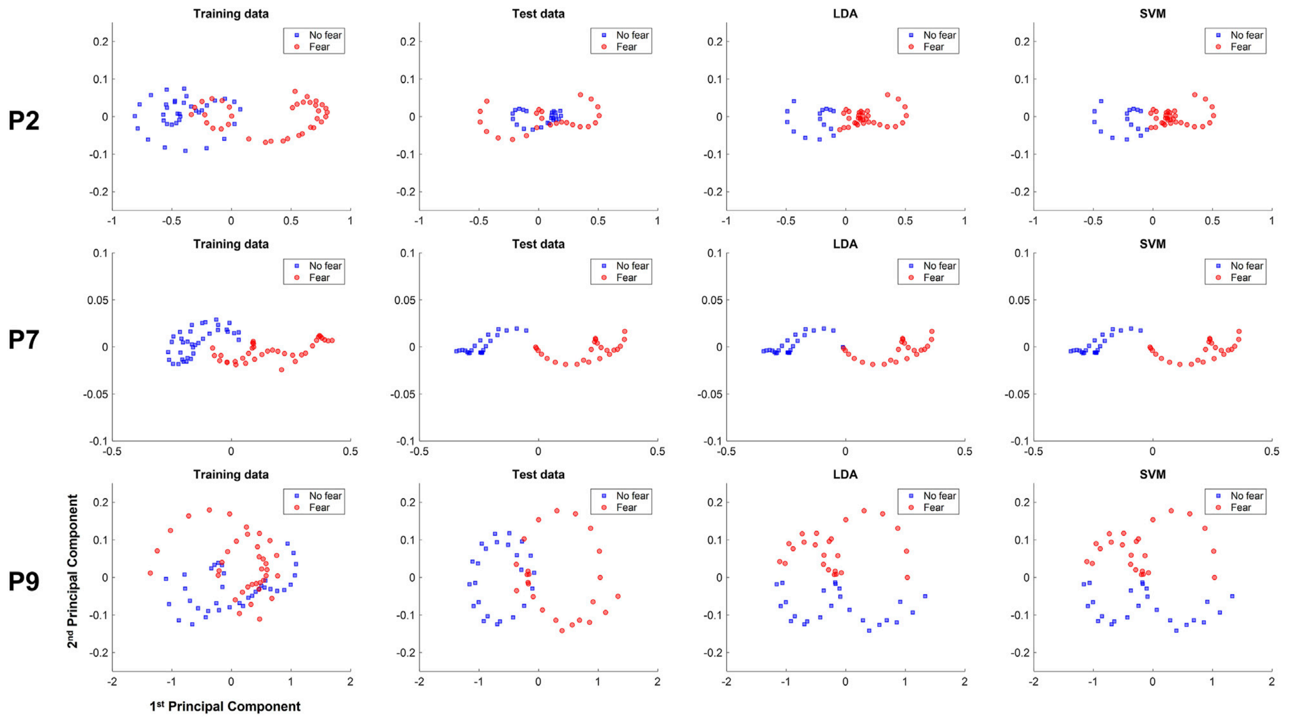


FIGURE 5 | Training, test data, and classification decisions from LDA and SVM of the 1-s subject-dependent classification of participants 2 (**P2**), 7 (**P7**), and 10 (**P10**), plotted against the first and second principle components.

TABLE 3 | Accuracies and confidence intervals [lower bound, upper bound] of the subject-independent classification).

Participant	1s history		3s history		5s history	
	LDA	SVM	LDA	SVM	LDA	SVM
1	78.33 (70.14,84.76)*	78.33 (70.14,84.76)*	75.83 (67.45,82.61)*	75.83 (67.45,82.61)*	76.67 (68.35,83.34)*	76.67 (68.35,83.34)*
2	67.50 (58.69,75.22)*	66.67 (57.83,74.47)*	70.83 (62.15,78.22)*	70.83 (62.15,78.22)*	70.83 (62.15,78.22)*	70.83 (62.15,78.22)*
3	79.17 (71.06,85.47)*	80.00 (71.96,86.18)*	78.33 (70.14,84.76)*	78.33 (70.14,84.76)*	77.50 (69.24,84.05)*	76.67 (68.35,83.34)*
4	84.17 (76.59,89.63)*	82.50 (74.72,88.26)*	89.17 (82.35,93.56)*	89.17 (82.35,93.56)*	88.33 (81.36,92.92)*	89.17 (82.35,93.56)*
5	85.00 (77.53,90.30)*	85.00 (77.53,90.30)*	90.83 (84.32,94.80)*	90.83 (84.32,94.80)*	89.17 (82.35,93.56)*	90.83 (84.32,94.80)*
6	72.50 (63.91,79.70)*	72.50 (63.91,79.70)*	84.17 (76.59,89.63)*	81.67 (73.80,87.57)*	82.50 (74.72,88.26)*	81.67 (73.80,87.57)*
7	88.33 (81.36,92.92)*	85.83 (78.48,90.96)*	90.00 (83.33,94.19)*	86.67 (79.44,91.63)*	91.67 (85.34,95.41)*	88.33 (81.36,92.92)*
8	48.33 (39.58,57.18)	51.67 (42.82,60.42)	50.00 (41.19,58.81)	50.00 (41.19,58.81)	52.50 (43.63,61.22)	54.17 (45.26,62.82)
9	65.83 (56.97,73.71)*	65.83 (56.97,73.71)*	68.33 (59.55,75.97)*	69.17 (60.42,76.73)*	67.50 (58.69,75.22)*	68.33 (59.55,75.97)*
10	87.50 (80.40,92.28)	85.83 (78.48,90.96)	91.67 (85.34,95.41)	88.33 (81.36,92.92)	90.83 (84.32,94.80)	89.17 (82.35,93.56)
11	53.33 (44.44,62.01)*	55.83 (46.90,64.40)*	55.00 (46.08,63.61)*	55.00 (46.08,63.)]*	54.17 (45.26,62.82)*	54.17 (45.26,62.82)*
12	68.52 (59.75,76.15)*	67.59 (58.78,75.31)*	63.89 (54.99,71.93)*	65.74 (56.88,73.62)*	60.19 (51.25,68.50)*	66.67 (57.83,74.47)*
13	69.17 (60.42,76.73)*	68.33 (59.55,75.97)*	69.17 (60.42,76.73)*	68.33 (59.55,75.97)*	70.00 (61.28,77.47)*	69.17 (60.42,76.73)*
14	57.50 (48.56,65.98)	57.50 (48.56,65.98)	63.33 (54.42,71.41)*	63.33 (54.42,71.41)*	65.83 (56.97,73.71)*	66.67 (57.83,74.47)*
Mean (±SD)	71.80 (69.60,73.90)* (±12.75)	71.67 (69.47,73.77)* (±11.60)	74.33 (72.19,76.36)* (±13.66)	73.80 (71.64,75.85)* (±12.84)	74.12 (71.97,76.16)* (±13.31)	74.46 (72.32,76.49)* (±12.34)

* *indicates the performance for which the lower bound of its confidence interval is larger than 50%.*

colors. The decisions made for the test data by the LDA and linear-SVM classifiers are also depicted. It should be noted that data distribution in feature space of 3- and 5-s history data are generally similar to 1-s history data and therefore not shown here. Also, it is worth mentioning that PCA is applied for visualization purposes only and not for feature dimension reduction.

In the data of participant 2 there is a high overlap in the test data between the distribution of the fear and non-fear classes, making it difficult to reach decent performance. In participant 9 data, LDA and linear-SVM learned to distinguish classes along the second PC, while the test data are clearly separable along the first PC, thereby yielding performance around chance level. In

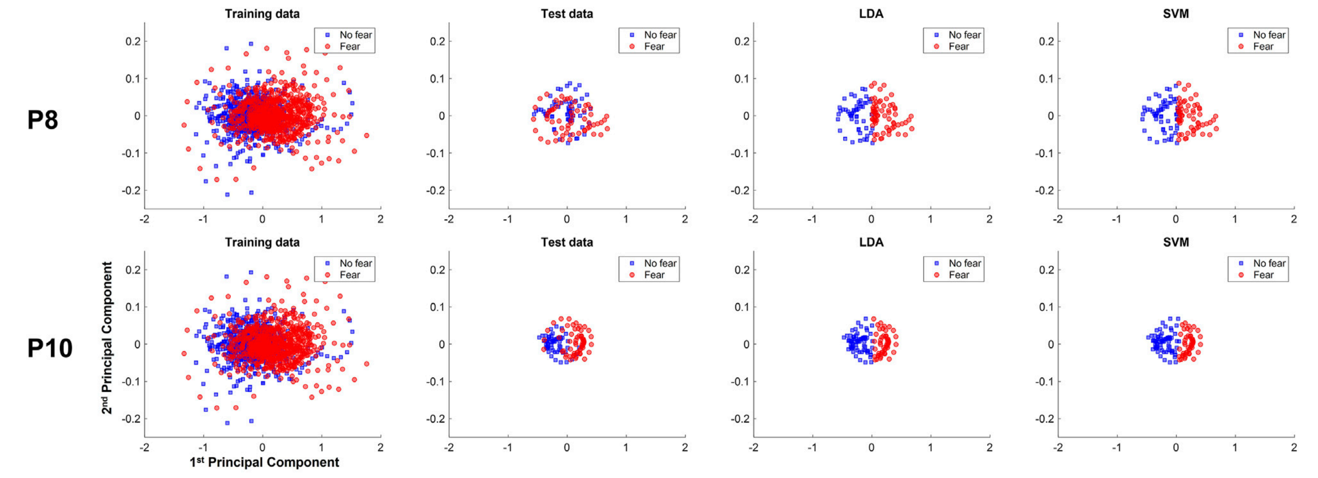


FIGURE 6 | Training, test data, and classification decisions from LDA and SVM of the 1-s subject-dependent classification of participants 2 (**P2**) and 10 (**P10**), plotted against the first and second principle components.

contrast, data distribution of training and test data are rather similar for participant 7, with a much clearer linear separability in test data. The classification for this participant is therefore high for of the linear classifiers.

3.3.2 Subject-independent Classification

Table 3 shows the accuracies of the subject-independent classifiers and the 95% confidence interval, calculated using **Eq. 1**. For the subject-independent classification, the 95% confidence intervals are smaller, since these are tested on $n = 120$ samples. This also implies that if the estimated performance is above 59% then the lower bound of the 95% confidence interval is larger than 50%. This implies that for most types of classifiers 12 out of 14 subject-independent classifiers perform significantly better than random. In order to compare the mean performance of the subject-independent classification with the mean of subject-independent random classification, we take the same approach as for the subject-dependent classification. In this case we have 14×120 trails, and the 99% confidence interval is given by (47%, 53%). Based on the mean accuracies computed from all participants, it can be inferred that the SVM on the 5-s history performs best, with a mean accuracy of 74.46% (SD 12.34). On the contrary, the SVM on the 1-s history performs the worst on average, with a mean accuracy of 71.67% (SD 11.60). From the calculated 99% confidence interval (47%, 53%) for the mean of random subject-independent classification (see **Section 2.6.2**), one can easily deduce that the mean of the subject-dependent classification is significantly higher (99.5% confidence) than the mean of random subject-independent classification.

Again, the difference between the accuracies of the classifiers that perform best and worst on average is only a few percent, indicating that the amount of history and the classification methods have merely minor influence on the classification performance. Again, the accuracies vary considerably among participants, ranging from 48.33% (participant 8) to 91.67% (participant 10), and the cause of this variation is also investigated by data distribution in PC space.

In the data of participant 8, the training data depicted in **Figure 6** (P8) shows that the data labeled as *no fear* are mostly centered around the negative values of the first PC, while *fear* data were located around the positive values. However, the test data of **Figure 6** (P8) have a different pattern. Instead, the data of the different labels are distributed over the positive and negative values of the second PC and are quite overlapping, indicating the difficulty to separate the test data of the different labels by a linear decision boundary. This might explain why the classifiers, which seemingly learned to separate classes along the first PC, cannot perform well on the test data, yielding low accuracy. In contrast, data distribution in test data from participant 10 [see **Figure 6** (P10)] are linearly separable in the first PC. Specifically, data labeled as *no fear* are centered around the negative values of the first PC, and the test data labeled as *fear* are centered around the positive values of the first PC. The linear classifiers were therefore successful in generalizing the learned pattern along the first PC to the test data, achieving high classification performance.

4 DISCUSSION

The aim of this study was to measure brain activity of participants with and without fear of heights when exposed to fearful stimuli presented in VR-HMD. Additionally, the study investigated the feasibility to train a simple classifier to recognize a fear response from fNIRS signals recorded from participants with fear of heights. A successful combination of fNIRS measurement and VR-HMD can prove the ecological validity of its use in VRET.

4.1 Statistical Analyses

4.1.1 Between-Group Analysis

The results from the between-group analysis of the fNIRS signals showed that the grand average contrast Δ[HbO] signals of the control group and the experimental group are significantly different in channel 3. No significant differences were found between the grand average contrast Δ[HbR] signals of the two groups. The evidence that only one out of 27 channels shows a significant difference between the two groups for only one chromophore suggests that the fNIRS signals of participants with fear of heights were not that different from those of participants without fear of heights in general.

However, it is difficult to make a direct comparison of our result with the literature due to the lack of including both experimental and control groups in previous studies on this topic. Despite this, it was discovered that Δ[HbO] measured in (some areas of) the PFC of recruited homogeneous participants increased during fearful conditions (Rosenbaum et al., 2020; Glotzbach et al., 2011; Zhang et al., 2017; Köchel et al., 2013; Roos et al., 2011; Ma et al., 2013; Landowska, 2018), which is in line with our results that the Δ[HbO] signal of the experimental group peaks higher than that of the control group when exposed to fearful stimuli (see **Figure 3**). In contrast, we found that Δ[HbR] of both groups were quite equal, which is partly in accordance with the evidence that the majority of similar works did not report any change of Δ[HbR] after the exposure to fearful stimuli (Roos et al., 2011; Ma et al., 2013; Zhang et al., 2017; Rosenbaum et al., 2020), but there are some exceptions (Glotzbach et al., 2011; Köchel et al., 2013; Landowska et al., 2018).

Nevertheless, it was also reported in the literature that Δ[HbO] values over the PFC can increase when the participants were experiencing other mental states, such as mental workload, mental stress, affective responses, attention, deception, preference, anticipation, suspicion, and frustration (Suzuki et al., 2008; Ayaz et al., 2012; Kreplin and Fairclough, 2013; Ding et al., 2014; Hirshfield et al., 20142014; Tupak et al., 2014; Arefi Shirvan et al., 2018; Numata et al., 2019). This indicates that increased Δ[HbO] values are not only an indication of fear responses but can also be driven by other psychological factors. This effect is less likely for the experimental group in our study, as they indicated that they were feeling afraid during the height exposure, which makes it improbable that they also experienced other mental states, considering that fear is presumably the most salient feeling they would perceive.

4.1.2 Within-Group Analysis

The result of the within-group analysis of the fNIRS signals shows that the grand average Δ[HbO] values are significantly higher during the height condition than during the ground condition. This significant difference was observed in a total of 13 channels, which are all located towards the frontal part of the PFC. These results indicate that during fear responses, the Δ[HbO] values increase significantly as compared to no-fear responses, which is in accordance with the vast majority of previous works on fNIRS measurements taken during fear responses (Glotzbach et al., 2011; Roos et al., 2011; Köchel et al., 2013; Ma et al., 2013; Zhang et al., 2017; Landowska, 2018; Rosenbaum et al., 2020).

Additionally, the results of the within-group analysis show that the grand average Δ[HbR] values of the height condition and the ground condition are significantly different in channel 23. Surprisingly, the grand average Δ[HbR] signal of the height condition is higher than that of the ground condition in this channel. This contradicts some findings from the literature, where decreased Δ[HbR] values are reported for fearful conditions (Glotzbach et al., 2011; Köchel et al., 2013; Landowska et al., 2018). It remains unclear why our results differ from the literature.

The clear distinction of fNIRS signals due to fear exposure in experimental group suggests the possibility to train a classifier to automatically detect fear responses using fNIRS.

4.2 Classification

4.2.1 Subject-Dependent Classification

The subject-dependent classification results suggest that the amount of history and the choice between the LDA or the linear-SVM algorithm has minimal influence on the subject-dependent classification performance. The linear classifiers do not perform well for some participants due to the difference in data distribution between training and test data. A possible explanation is that the fear responses and accompanying fNIRS measurements of these participants were not stable over time.

4.2.2 Subject-Independent Classification

Similarly, the choice of classification methods and the amount of history to take into account do not have enormous influence on the performance of subject-independent classifiers. While the overall accuracy is above 71%, classification for participants 8 and 11 achieved poor performance. Our investigation on the participants' AQ, SUDS, and IPQ scores indicated that these participants had a strong fear of heights, felt very anxious during the height trials, felt relaxed during the ground trials, and felt sufficiently present in the VEs. However, we learned from the data distribution analysis that the fNIRS measurements of these participants were not stable over time. This might explain the overlap of *fear* and *no fear* fNIRS data trials in the first two PCs of the feature space when taking all data from this participant as a test set (in leave-one-subject-out cross-validation), while the cause of the instability of data over time remains unclear.

It is remarkable that for most participants, the subject-independent classification outperforms the subject-dependent classification. While the training data from the first six trials have a different distribution than the last four trials used as a test set in subject-dependent classification, combining all trials might mitigate the discrepancy between those trials, converging to more common patterns of the other participants. This would explain the relatively good performance of the subject-independent classification. More research is needed to prove this hypothesis.

4.2.3 Overall Classification Performance

Overall, the average classification accuracy of the subject-dependent classification is approximately 70%, whereas the subject-independent classification has average accuracies around 74%. These accuracies are statistically significantly higher than random classification. It is noteworthy that the goal of this study is not to find the best classification model but to examine to what extent a simple linear classifier with minimal parameter tuning can discriminate between fear and no-fear responses. Future work on applying sophisticated algorithms could improve the classification performance. It should also be noted that our channel selection used in classification was based solely on signal quality and not influenced by feature correlation.

The accuracies in our study are comparable to those in previous works attempting to classify fear from no-fear responses in VR using physiological signals. Despite achieving slightly higher accuracies from 76% to 89.5% compared to our research, previous work recruited a lower number of participants [seven participants in Handouzi et al. (20132013) using blood volume pulse (BVP) data, eight participants in Bălan et al. (2020) using galvanic skin response (GSR), heart rate, and EEG data]. Among similar studies that recruited a higher number of participants, the study by Šalkevicius et al. (2019) detected public speaking anxiety using BVP, GSR, and skin temperature data from 30 participant and achieved 80.1% accuracy in leave-one-subject-out cross-validation. However, this work did not include brain signals in the study. In contrast, another study (Hu et al., 2018) detecting fear of heights response in VR-HMD from EEG signal achieved 88.77%, but the results were based on 10-fold cross-validation, where the generalizability to classify unseen participant remains unknown.

In general, our classification performance has demonstrated the feasibility to detect a fear of height response from brain signals of a previously unseen participant as a 1-s rate (the size of our sliding window is 1 s). As we encourage other researchers to test other classification paradigms, the physiological data reproducing the results in this study are publicly available.

4.3 Limitations of the Study

It is noteworthy that the fNIRS signals comprise multiple components where some of them are potentially confounders that are not task-related. Our method is based on contrasting an experimental condition with a baseline condition, which can subtract out spurious hemodynamic/oxygenation responses from the experimental task (Tachtsidis and Scholkmann, 2016). Thus, it should reduce hemodynamic influences from the extracerebral layer from the fNIRS signals. Alternative approaches can be further incorporated to remove systemic confounders.

Neither the fNIRS headset nor the VR-HMD was originally designed for simultaneous usage of both devices. The incompatibility caused an uncomfortable feeling for many participants. The VR-HMD needed tightening up with a headband around the participant's head. This put an extra pressure on the optodes of the fNIRS headset, which can be unpleasant for some participants, negatively influencing the user experience of the system. Since it is difficult to quantify the effect of the uncomfortable feeling caused by the hardware components, it remains unknown to what extent this affected the participants and the consequent results.

Although the fNIRS technology is less susceptible to motion artifacts and electrical noise, it was often reported in the literature that motion artifacts still occur (Naseer and Hong, 2015; Wilcox and Biondi, 2015; Pinti et al., 2020). Therefore, the participants in our study were instructed to look around very slowly in the VEs and to limit their bodily movements, which might reduce realism of the experience of the VEs for some participants. Still, the motion artifacts were present in our study.

4.4 Recommendations for Future Work

Future research should consider adding more trials per condition and prolonging the duration of each trial. More data are needed to train sophisticated classification models. Prolonged duration opens the possibility to include heart rate variability (HRV) as a feature for the classifier; HRV can be captured from the embedded cardiac cycles in the fNIRS signals and was found to be a useful measure to detect fear (Wiederhold et al., 2002; Peterson et al., 2018). Also, it is worthy to investigate the difference among high-arousal-negative-valence responses, such as mental stress, frustration, and fear; disentanglement of such responses can potentially improve the detection of fear of height.

In this research, we measure distress and the feeling of presence by using established scales to allow the comparison with previous research in fear and VRET. While SUDS has been widely used in the context of fear exposure treatment due to its high comprehensiveness, conciseness, and validity in psychological studies, future works should also consider using recently developed measures that are correlated highly with the SUDS to confirm the validity of the measured fear by the classical SUDS. These include the scale of anxiety (Spielberger, 1972; Masia-Warner et al., 2003), discomfort (Kaplan et al., 1995), disturbance (Harris et al., 2002; Kim et al., 2008), or distress (McCullough, 2002). Similarly, although IPQ was found as the most reliable questionnaire to measure the presence in VR environment (Schwind et al., 2019) among classical measures (Witmer and Singer, 1998; Slater and Steed, 2000; Usoh et al., 2000), the alternative recent questionnaires should also be considered (Grassini and Laumann, 2020).

The statistical analyses of the fNIRS data and the classification performances are merely based on the mean Δ[HbO] and mean Δ[HbR] values. However, it is known from previous fNIRS studies investigating mental states that alternative features such as amplitude, slope, standard deviation, kurtosis, skewness, and signal peaks can provide insights and be used as discriminative features for classifying mental states (Khan and Hong, 2015; Zhang et al., 2016; Aghajani et al., 2017; Parent et al., 2019). In our study, the grand average Δ[HbO] traces revealed that the traces of the experimental group generally rise to a peak value, whereas this pattern is less apparent for the grand average traces of the control group (see **Figure 3**). A similar observation can be made for the grand average Δ[HbO] traces of the height condition and ground condition of the experimental group (see **Figure 4**). Based on these observations, it is anticipated that alternative features such as the maximum signal value, the time to peak, and the signal slope have the potential to improve the classification results or enhance the fNIRS difference between groups.

5 CONCLUSION

The results answer our first research question by demonstrating that there is significant difference in fNIRS signals between participants with a fear of heights and participants without it when exposed to fear conditions in a VE. Specifically, the contrast between the ground-condition and height-condition fNIRS signals in the experimental group was larger than that in the control group, despite limited statistical significance. The effect of the condition was more salient when focusing only on the experimental group that exhibited significant differences in the grand average Δ[HbO] values during fear responses and during no-fear responses. The effect was dominant in the optode area close to the frontal part of the PFC. To answer our research question regarding to what extent a machine learning model can be successfully trained to recognize fear of heights response using fNIRS, we trained different simple classifiers in a subject-dependent and subject-independent framework and found that subject-dependent classification encountered the issue of subjective variability. Nevertheless, the subject-independent classification results show the potential for usage in online fear of height detection, and the average accuracy in classifying unseen data from a previously unseen participant is above 74.00%. Our study therefore confirmed the ecological validity of combining fNIRS measurement and VR-HMD, which may pave a way toward effective VRET.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: [https://doi.org/10.4121/17302865].

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente (reference number: RP 2020-76). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LdW, NT, and MP conceived, planned, designed the experiment. LdW provided a critical review, carried out the experiment, data acquisition, analysis, processing interpretation, and wrote the full research report. NT wrote the manuscript with input from all the authors. MP supervised the research, provided critical feedback, and proofread the manuscript. All authors discussed the results and contributed to the final manuscript.

FUNDING

This work was partially supported by the European Regional Development Fund's operationeel programma oost (OP-OOST EFRO PROJ-00900) and by the Netherlands Organization for Scientific Research (NWA Startimpuls 400.17.602).

ACKNOWLEDGMENTS

The authors gratefully acknowledge Tenzing Dolmans, faculty members, and staff at the Faculty of Behavioural, Management and Social sciences, University of Twente (UT) for providing lab space and materials required for the experiments and facilitating the study. The authors also thank the participants for their time and effort in participating in the experiment. The authors also express gratitude to Dirk Heylen at Human-Media Interaction group, UT, for advice and discussion.

REFERENCES

Aghajani, H., Garbey, M., and Omurtag, A. (2017). Measuring Mental Workload with EEG+fNIRS. *Front. Hum. Neurosci.* 11. doi:10.3389/fnhum.2017.00359

Aksoy, E., Izzetoglu, K., Baysoy, E., Agrali, A., Kitapcioglu, D., and Onaral, B. (2019). Performance Monitoring via Functional Near Infrared Spectroscopy for Virtual Reality Based Basic Life Support Training. *Front. Neurosci.* 13. doi:10.3389/fnins.2019.01336

Antony, M. M. (2001). *Measures for Specific Phobia*. Boston, MA: Springer US, 133–158. doi:10.1007/0-306-47628-2_12

Arefi Shirvan, R., Setarehdan, S. K., and Motie Nasrabadi, A. (2018). Classification of Mental Stress Levels by Analyzing fNIRS Signal Using Linear and Non-linear Features. *Int. Clin. Neurosci. J.* 5, 55–61. doi:10.15171/icnj.2018.11

Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., and Onaral, B. (2012). Optical Brain Monitoring for Operator Training and Mental Workload Assessment. *NeuroImage* 59, 36–47. Neuroergonomics: The human brain in action and at work. doi:10.1016/j.neuroimage.2011.06.023

Bălan, O., Moise, G., Moldoveanu, A., Leordeanu, M., and Moldoveanu, F. (2020). An Investigation of Various Machine and Deep Learning Techniques Applied in Automatic Fear Level Detection and Acrophobia Virtual Therapy. *Sensors* 20, 496. doi:10.3390/s20020496

Benjamin, C. L., O'Neil, K. A., Crawley, S. A., Beidas, R. S., Coles, M., and Kendall, P. C. (2010). Patterns and Predictors of Subjective Units of Distress in Anxious Youth. *Behav. Cogn. Psychother.* 38, 497–504. doi:10.1017/S1352465810000287

Boeldt, D., McMahon, E., McFaul, M., and Greenleaf, W. (2019). Using Virtual Reality Exposure Therapy to Enhance Treatment of Anxiety Disorders: Identifying Areas of Clinical Adoption and Potential Obstacles. *Front. Psychiatry* 10. doi:10.3389/fpsyt.2019.00773

Brinkman, W.-P., Sandino, G., and Mast, C. (2009). "Field Observations of Therapists Conducting Virtual Reality Exposure Treatment for the Fear of Flying," in *VTT Symposium (Valtion Teknillinen Tutkimuskeskus)*.

Chen, L.-C. (2016). *Cortical Plasticity in Cochlear Implant Users*. Ph.D. thesis. doi:10.13140/RG.2.2.33251.76324

Cohen, D. C. (1977). Comparison of Self-Report and Overt-Behavioral Procedures for Assessing Acrophobia. *Behav. Ther.* 8, 17–23. doi:10.1016/S0005-7894(77)80116-0

Cui, X., Bray, S., and Reiss, A. L. (2010). Functional Near Infrared Spectroscopy (NIRS) Signal Improvement Based on Negative Correlation between Oxygenated and Deoxygenated Hemoglobin Dynamics. *NeuroImage* 49, 3039–3046. doi:10.1016/j.neuroimage.2009.11.050

Delpy, D. T., Cope, M., Zee, P. v. d., Arridge, S., Wray, S., and Wyatt, J. (1988). Estimation of Optical Pathlength through Tissue from Direct Time of Flight Measurement. *Phys. Med. Biol.* 33, 1433–1442. doi:10.1088/0031-9155/33/12/008

Derosière, G., Dalhoumi, S., Perrey, S., Dray, G., and Ward, T. (2014). Towards a Near Infrared Spectroscopy-Based Estimation of Operator Attentional State. *PLOS ONE* 9, e92045. doi:10.1371/journal.pone.0092045

Difede, J., and Hoffman, H. G. (2002). Virtual Reality Exposure Therapy for World Trade center post-traumatic Stress Disorder: A Case Report. *CyberPsychology Behav.* 5, 529–535. doi:10.1089/109493102321018169

Ding, X. P., Sai, L., Fu, G., Liu, J., and Lee, K. (2014). Neural Correlates of Second-Order Verbal Deception: A Functional Near-Infrared Spectroscopy (fNIRS) Study. *NeuroImage* 87, 505–514. doi:10.1016/j.neuroimage.2013.10.023

Dong, D., Wong, L. K. F., and Luo, Z. (2018). Assess Ba10 Activity in Slide-Based and Immersive Virtual Reality Prospective Memory Task Using Functional Near-Infrared Spectroscopy (fNIRS). *Appl. Neuropsychol. Adult* 26, 465–471. doi:10.1080/23279095.2018.1443104

Dong, D., Wong, L. K. F., and Luo, Z. (2017). Assessment of Prospective Memory Using fNIRS in Immersive Virtual Reality Environment. *Jbbs* 07, 247–258. doi:10.4236/jbbs.2017.76018

Donker, T., Van Esveld, S., Fischer, N., and Van Straten, A. (2018). 0Phobia - towards a Virtual Cure for Acrophobia: Study Protocol for a Randomized Controlled Trial. *Trials* 19. doi:10.1186/s13063-018-2704-6

Emmelkamp, P. M. G., Bruynzeel, M., Drost, L., and van der Mast, C. A. P. G. (2001). Virtual Reality Treatment in Acrophobia: A Comparison with Exposure *In Vivo*. *CyberPsychology Behav.* 4, 335–339. doi:10.1089/109493101300210222

Etkin, A., and Wager, T. D. (2007). Functional Neuroimaging of Anxiety: A Meta-Analysis of Emotional Processing in Ptsd, Social Anxiety Disorder, and Specific Phobia. *Ajp* 164, 1476–1488. doi:10.1176/appi.ajp.2007.07030504

Fishburn, F. A., Ludlum, R. S., Vaidya, C. J., and Medvedev, A. V. (2019). Temporal Derivative Distribution Repair (Tddr): A Motion Correction Method for fNIRS. *NeuroImage* 184, 171–179. doi:10.1016/j.neuroimage.2018.09.025

Freeman, D., Haselton, P., Freeman, J., Spanlang, B., Kishore, S., Albery, E., et al. (2018). Automated Psychological Therapy Using Immersive Virtual Reality for Treatment of Fear of Heights: a Single-Blind, Parallel-Group, Randomised Controlled Trial. *The Lancet Psychiatry* 5, 625–632. doi:10.1016/S2215-0366(18)30226-8

Garcia-Palacios, A., Hoffman, H., Carlin, A., Furness, T. A., and Botella, C. (2002). Virtual Reality in the Treatment of Spider Phobia: a Controlled Study. *Behav. Res. Ther.* 40, 983–993. doi:10.1016/S0005-7967(01)00068-7

Genovese, C. R., Lazar, N. A., and Nichols, T. (2002). Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate. *NeuroImage* 15, 870–878. doi:10.1006/nimg.2001.1037

Gerardi, M., Rothbaum, B. O., Ressler, K., Heekin, M., and Rizzo, A. (2008). Virtual Reality Exposure Therapy Using a Virtual iraq: Case Report. *J. Traum. Stress* 21, 209–213. doi:10.1002/jts.20331

Glotzbach, E., Mühlberger, A., Gschwendtner, K., Fallgatter, A. J., Pauli, P., and Herrmann, M. J. (2011). Prefrontal Brain Activation during Emotional Processing: A Functional Near Infrared Spectroscopy Study (fNIRS). *Open Neuroimag J.* 5, 33–39. doi:10.2174/1874440001105010033

Grassini, S., and Laumann, K. (2020). Questionnaire Measures and Physiological Correlates of Presence: A Systematic Review. *Front. Psychol.* 11, 349. doi:10.3389/fpsyg.2020.00349

Gromer, D., Madeira, O., Gast, P., Nehfischer, M., Jost, M., Müller, M., et al. (2018). Height Simulation in a Virtual Reality Cave System: Validity of Fear Responses and Effects of an Immersion Manipulation. *Front. Hum. Neurosci.* 12. doi:10.3389/fnhum.2018.00372

Handouzi, W., Maaoui, C., Pruski, A., Moussaoui, A., and Bendiouis, Y. (20132013). Short-term Anxiety Recognition Induced by Virtual Reality Exposure for Phobic People. *IEEE Int. Conf. Syst. Man, Cybernetics*, 3145–3150. doi:10.1109/smc.2013.536

Harris, S. R., Kemmerling, R. L., and North, M. M. (2002). Brief Virtual Reality Therapy for Public Speaking Anxiety. *Cyberpsychol Behav.* 5, 543–550. doi:10.1089/109493102321018187

Hill, A. P., and Bohil, C. J. (2016). Applications of Optical Neuroimaging in Usability Research. *Ergon. Des.* 24, 4–9. doi:10.1177/1064804616629309

Hirshfield, L. M., Bobko, P., Barelka, A., Hirshfield, S. H., Farrington, M. T., Gulbronson, S., et al. (20142014). Using Noninvasive Brain Measurement to Explore the Psychological Effects of Computer Malfunctions on Users during Human-Computer Interactions. *Adv. Human-Computer Interaction* 2014, 1–13. doi:10.1155/2014/101038

Hocke, L., Oni, I., Duszynski, C., Corrigan, A., FrederickFrederick, B., and Dunn, J. (2018). Automated Processing of fNIRS Data-A Visual Guide to the Pitfalls and Consequences. *Algorithms* 11, 67–92. doi:10.3390/a11050067

Hofmann, S. (2008). Cognitive Processes during Fear Acquisition and Extinction in Animals and Humans: Implications for Exposure Therapy of Anxiety Disorders. *Clin. Psychol. Rev.* 28, 199–210. doi:10.1016/j.cpr.2007.04.009

Hu, F., Wang, H., Chen, J., and Gong, J. (2018). "Research on the Characteristics of Acrophobia in Virtual Altitude Environment," in 2018 IEEE International Conference on Intelligence and Safety for Robotics (ISR), 238–243. doi:10.1109/iisr.2018.8535774

Hu, X.-S., Hong, K.-S., and Ge, S. S. (2012). fNIRS-Based Online Deception Decoding. *J. Neural Eng.* 9, 026012. doi:10.1088/1741-2560/9/2/026012

Hudak, J., Blume, F., Dresler, T., Haeussinger, F. B., Renner, T. J., Fallgatter, A. J., et al. (2017). Near-infrared Spectroscopy-Based Frontal Lobe Neurofeedback Integrated in Virtual Reality Modulates Brain and Behavior in Highly Impulsive Adults. *Front. Hum. Neurosci.* 11. doi:10.3389/fnhum.2017.00425

Jerdan, S. W., Grindle, M., van Woerden, H. C., and Kamel Boulos, M. N. (2018). Head-mounted Virtual Reality and Mental Health: Critical Review of Current Research. *JMIR Serious Games* 6, e14. doi:10.2196/games.9226

Kaplan, D. M., Smith, T., and Coons, J. (1995). A Validity Study of the Subjective Unit of Discomfort (SUD) Score. Measurement and Evaluation in Counseling and Development 27.

Khan, M. J., and Hong, K.-S. (2015). Passive Bci Based on Drowsiness Detection: An fNIRS Study. *Biomed. Opt. Express* 6, 4063–4078. doi:10.1364/BOE.6.004063

Khan, M. N. A., Bhutta, M. R., and Hong, K.-S. (2020). Task-specific Stimulation Duration for fNIRS Brain-Computer Interface. *IEEE Access* 8, 89093–89105. doi:10.1109/access.2020.2993620

Kim, D., Bae, H., and Chon Park, Y. (2008). Validity of the Subjective Units of Disturbance Scale in EMDR. *J. EMDR Prac Res.* 2, 57–62. doi:10.1891/1933-3196.2.1.57

Kim, G., Buntain, N., Hirshfield, L., Costa, M. R., and Chock, T. M. (2019). *Processing Racial Stereotypes in Virtual Reality: An Exploratory Study Using Functional Near-Infrared Spectroscopy (fNIRS)*. Springer Cham, 407–417. doi:10.1007/978-3-030-22419-6_29

Köchel, A., Schöngassner, F., and Schienle, A. (2013). Cortical Activation during Auditory Elicitation of Fear and Disgust: A Near-Infrared Spectroscopy (NIRS) Study. *Neurosci. Lett.* 549, 197–200. doi:10.1016/j.neulet.2013.06.062

Kreplin, U., and Fairclough, S. H. (2013). Activation of the Rostromedial Prefrontal Cortex during the Experience of Positive Emotion in the Context of Esthetic Experience. An fNIRS Study. *Front. Hum. Neurosci.* 7. doi:10.3389/fnhum.2013.00879

Lachert, P., Janusek, D., Pulawski, P., Liebert, A., Milej, D., and Blinowska, K. J. (2017). Coupling of Oxy- and Deoxyhemoglobin Concentrations with EEG Rhythms during Motor Task. *Sci. Rep.* 7, 15414. doi:10.1038/s41598-017-15770-2

Landowska, A. (2018). "Measuring Prefrontal Cortex Response to Virtual Reality Exposure Therapy,". Ph.D. thesis (Salford, United Kingdom: University of Salford).*Freely Moving Participants.*

Landowska, A., Roberts, D., Eachus, P., and Barrett, A. (2018). Within- and Between-Session Prefrontal Cortex Response to Virtual Reality Exposure Therapy for Acrophobia. *Front. Hum. Neurosci.* 12. doi:10.3389/fnhum.2018.00362

Lange, K., Williams, L. M., Young, A. W., Bullmore, E. T., Brammer, M. J., Williams, S. C. R., et al. (2003). Task Instructions Modulate Neural Responses to Fearful Facial Expressions. *Biol. Psychiatry* 53, 226–232. doi:10.1016/S0006-3223(02)01455-5

Lindner, P., Miloff, A., Bergman, C., Andersson, G., Hamilton, W., and Carlbring, P. (2020). Gamified, Automated Virtual Reality Exposure Therapy for Fear of Spiders: A Single-Subject Trial under Simulated Real-World Conditions. *Front. Psychiatry* 11, 116. doi:10.3389/fpsyt.2020.00116

Ma, Q., Huang, Y., and Wang, L. (2013). Left Prefrontal Activity Reflects the Ability of Vicarious Fear Learning: A Functional Near-Infrared Spectroscopy Study. *Scientific World J.* 2013, 1–8. doi:10.1155/2013/652542

Maltby, N., Kirsch, I., Mayers, M., and Allen, G. J. (2002). Virtual Reality Exposure Therapy for the Treatment of Fear of Flying: a Controlled Investigation. *J. Consulting Clin. Psychol.* 70, 1112–1118. doi:10.1037//0022-006x.70.5.111210.1037/0022-006x.70.5.1112

Martens, M. A., Antley, A., Freeman, D., Slater, M., Harrison, P. J., and Tunbridge, E. M. (2019). It Feels Real: Physiological Responses to a Stressful Virtual Reality Environment and its Impact on Working Memory. *J. Psychopharmacol.* 33, 1264–1273. doi:10.1177/0269881119860156

Masia-Warner, C., Storch, E. A., Pincus, D. B., Klein, R. G., Heimberg, R. G., and Liebowitz, M. R. (2003). The Liebowitz Social Anxiety Scale for Children and Adolescents: An Initial Psychometric Investigation. *J. Am. Acad. Child Adolesc. Psychiatry* 42, 1076–1084. doi:10.1097/01.CHI.0000070249.24125.89

McCullough, L. (2002). Exploring Change Mechanisms in EMDR Applied to "Small-T Trauma" in Short-Term Dynamic Psychotherapy: Research Questions and Speculations. *J. Clin. Psychol.* 58, 1531–1544. doi:10.1002/jclp.10103

Miloff, A., Lindner, P., Dafgård, P., Deak, S., Garke, M., Hamilton, W., et al. (2019). Automated Virtual Reality Exposure Therapy for Spider Phobia vs. *In-Vivo* One-Session Treatment: A Randomized Non-inferiority Trial. *Behav. Res. Ther.* 118, 130–140. doi:10.1016/j.brat.2019.04.004

Naseer, N., and Hong, K.-S. (2015). fNIRS-Based Brain-Computer Interfaces: a Review. *Front. Hum. Neurosci.* 9, 3. doi:10.3389/fnhum.2015.00003

Nomura, M., Ohira, H., Haneda, K., Iidaka, T., Sadato, N., Okada, T., et al. (2004). Functional Association of the Amygdala and Ventral Prefrontal Cortex during Cognitive Evaluation of Facial Expressions Primed by Masked Angry Faces: an Event-Related Fmri Study. *NeuroImage* 21, 352–363. doi:10.1016/j.neuroimage.2003.09.021

Numata, T., Kiguchi, M., and Sato, H. (2019). Multiple-time-scale Analysis of Attention as Revealed by EEG, NIRS, and Pupil Diameter Signals during a Free Recall Task: A Multimodal Measurement Approach. *Front. Neurosci.* 13. doi:10.3389/fnins.2019.01307

Parent, M., Peysakhovich, V., Mandrick, K., Tremblay, S., and Causse, M. (2019). The Diagnosticity of Psychophysiological Signatures: Can We Disentangle Mental Workload from Acute Stress with ECG and fNIRS? *Int. J. Psychophysiology* 146, 139–147. doi:10.1016/j.ijpsycho.2019.09.005

Peterson, S. M., Furuichi, E., and Ferris, D. P. (2018). Effects of Virtual Reality High Heights Exposure during Beam-Walking on Physiological Stress and Cognitive Loading. *PLOS ONE* 13, e0200306. doi:10.1371/journal.pone.0200306

Pinti, P., Scholkmann, F., Hamilton, A., Burgess, P., and Tachtsidis, I. (2019). Current Status and Issues Regarding Pre-processing of fNIRS Neuroimaging Data: An Investigation of Diverse Signal Filtering Methods within a General Linear Model Framework. *Front. Hum. Neurosci.* 12, 505. doi:10.3389/fnhum.2018.00505

Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., et al. (2020). The Present and Future Use of Functional Near-infrared Spectroscopy (fNIRS) for Cognitive Neuroscience. *Ann. N.Y. Acad. Sci.* 1464, 5–29. doi:10.1111/nyas.13948

Price, R. B., Eldreth, D. A., and Mohlman, J. (2011). Deficient Prefrontal Attentional Control in Late-Life Generalized Anxiety Disorder: an Fmri Investigation. *Transl Psychiatry* 1, e46. doi:10.1038/tp.2011.46

Riva, G., Mantovani, F., Capideville, C. S., Preziosa, A., Morganti, F., Villani, D., et al. (2007). Affective Interactions Using Virtual Reality: The Link between Presence and Emotions. *CyberPsychology Behav.* 10, 45–56. doi:10.1089/cpb.2006.9993

Rizzo, A., John, B., Newman, B., Williams, J., Hartholt, A., Lethin, C., et al. (2013). Virtual Reality as a Tool for Delivering Ptsd Exposure Therapy and Stress Resilience Training. *Mil. Behav. Health* 1, 52–58. doi:10.1080/21635781.2012.721064

Rodríguez, A., Rey, B., Clemente, M., Wrzesien, M., and Alcañiz, M. (2015). Assessing Brain Activations Associated with Emotional Regulation during Virtual Reality Mood Induction Procedures. *Expert Syst. Appl.* 42, 1699–1709. doi:10.1016/j.eswa.2014.10.006

Roos, A., Robertson, F., Lochner, C., Vythilingum, B., and Stein, D. J. (2011). Altered Prefrontal Cortical Function during Processing of Fear-Relevant Stimuli in Pregnancy. *Behav. Brain Res.* 222, 200–205. doi:10.1016/j.bbr.2011.03.055

Rosenbaum, D., Leehr, E. J., Rubel, J., Maier, M. J., Pagliaro, V., Deutsch, K., et al. (2020). Cortical Oxygenation during Exposure Therapy - *In Situ* fNIRS Measurements in Arachnophobia. *NeuroImage: Clin.* 26, 102219. doi:10.1016/j.nicl.2020.102219

Rothbaum, B. O., Anderson, P., Zimand, E., Hodges, L., Lang, D., and Wilson, J. (2006). Virtual Reality Exposure Therapy and Standard (*In Vivo*) Exposure Therapy in the Treatment of Fear of Flying. *Behav. Ther.* 37, 80–90. doi:10.1016/j.beth.2005.04.004

Rothbaum, B. O., Hodges, L. F., Ready, D., Graap, K., and Alarcon, R. D. (2001). Virtual Reality Exposure Therapy for Vietnam Veterans with Posttraumatic Stress Disorder. *J. Clin. Psychiatry* 62, 617–622. doi:10.4088/JCP.v62n0808

Rothbaum, B. O., Hodges, L., Smith, S., Lee, J. H., and Price, L. (2000). A Controlled Study of Virtual Reality Exposure Therapy for the Fear of Flying. *J. Consulting Clin. Psychol.* 68, 1020–1026. doi:10.1037/0022-006X.68.6.1020

Rothbaum, B. O., Price, M., Jovanovic, T., Norrholm, S. D., Gerardi, M., Dunlop, B., et al. (2014). A Randomized, Double-Blind Evaluation Ofd-Cycloserine or Alprazolam Combined with Virtual Reality Exposure Therapy for Posttraumatic Stress Disorder in Iraq and Afghanistan War Veterans. *Ajp* 171, 640–648. doi:10.1176/appi.ajp.2014.13121625

Šalkevicius, J., Damaševičius, R., Maskeliunas, R., and Laukienė, I. (2019). Anxiety Level Recognition for Virtual Reality Therapy System Using Physiological Signals. *Electronics* 8, 1039. doi:10.3390/electronics8091039

Scholkmann, F., Kleiser, S., Metz, A. J., Zimmermann, R., Mata Pavia, J., Wolf, U., et al. (2014). A Review on Continuous Wave Functional Near-Infrared Spectroscopy and Imaging Instrumentation and Methodology. *NeuroImage* 85, 6–27. doi:10.1016/j.neuroimage.2013.05.004

Schubert, T., Friedmann, F., and Regenbrecht, H. (2001). The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators & Virtual Environments* 10, 266–281. doi:10.1162/105474601300343603

Schwind, V., Knierim, P., Haas, N., and Henze, N. (2019). *Using Presence Questionnaires in Virtual Reality*, 1–12. New York, NY, USA: Association for Computing Machinery.

Seraglia, B., Gamberini, L., Priftis, K., Scatturin, P., Martinelli, M., and Cutini, S. (2011). An Exploratory fNIRS Study with Immersive Virtual Reality: a New Method for Technical Implementation. *Front. Hum. Neurosci.* 5. doi:10.3389/fnhum.2011.00176

Shin, L. M., and Liberzon, I. (2010). The Neurocircuitry of Fear, Stress, and Anxiety Disorders. *Neuropsychopharmacol* 35, 169–191. doi:10.1038/npp.2009.83

Slater, M., and Steed, A. (2000). A Virtual Presence Counter. *Presence: Teleoperators & Virtual Environments* 9, 413–434. doi:10.1162/105474600566925

Spielberger, C. D. (1972). Review of State-Trait Anxiety Inventory. *Prof. Psychol.* 3.

Suzuki, M., Miyai, I., Ono, T., and Kubota, K. (2008). Activities in the Frontal Cortex and Gait Performance Are Modulated by Preparation. An fNIRS Study. *NeuroImage* 39, 600–607. doi:10.1016/j.neuroimage.2007.08.044

Tachtsidis, I., and Scholkmann, F. (2016). False Positives and False Negatives in Functional Near-Infrared Spectroscopy: Issues, Challenges, and the Way Forward. *Neurophoton* 3, 031405. doi:10.1117/1.NPh.3.3.031405

Teo, W.-P., Muthalib, M., Yamin, S., Hendy, A. M., Bramstedt, K., Kotsopoulos, E., et al. (2016). Does a Combination of Virtual Reality, Neuromodulation and Neuroimaging Provide a Comprehensive Platform for Neurorehabilitation? - A Narrative Review of the Literature. *Front. Hum. Neurosci.* 10. doi:10.3389/fnhum.2016.00284

Tupak, S. V., Dresler, T., Guhn, A., Ehlis, A.-C., Fallgatter, A. J., Pauli, P., et al. (2014). Implicit Emotion Regulation in the Presence of Threat: Neural and Autonomic Correlates. *NeuroImage* 85, 372–379. doi:10.1016/j.neuroimage.2013.09.066

Usoh, M., Catena, E., Arman, S., and Slater, M. (2000). Using Presence Questionnaires in Reality. *Presence: Teleoperators & Virtual Environments* 9, 497–503. doi:10.1162/105474600566989

Wald, J., and Taylor, S. (2000). Efficacy of Virtual Reality Exposure Therapy to Treat Driving Phobia: a Case Report. *J. Behav. Ther. Exp. Psychiatry* 31, 249–257. doi:10.1016/S0005-7916(01)00009-X

Wiederhold, B. K., Jang, D. P., Kim, S. I., and Wiederhold, M. D. (2002). Physiological Monitoring as an Objective Tool in Virtual Reality Therapy. *CyberPsychology Behav.* 5, 77–82. doi:10.1089/109493102753685908

Wilcox, T., and Biondi, M. (2015). fNIRS in the Developmental Sciences. *Wires Cogn. Sci.* 6, 263–283. doi:10.1002/wcs.1343

Wilson, E. B. (1927). Probable Inference, the Law of Succession, and Statistical Inference. *J. Am. Stat. Assoc.* 22, 209–212. doi:10.1080/01621459.1927.10502953

Witmer, B. G., and Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence* 7, 225–240. doi:10.1162/105474698565686

Wolpe, J. (1969). *The Practice of Behavior Therapy*. New York: Pergamon Press.

Zhang, D., Zhou, Y., Hou, X., Cui, Y., and Zhou, C. (2017). Discrimination of Emotional Prosodies in Human Neonates: A Pilot fNIRS Study. *Neurosci. Lett.* 658, 62–66. doi:10.1016/j.neulet.2017.08.047

Zhang, Z., Jiao, X., Jiang, J., Pan, J., Cao, Y., Yang, H., et al. (2016). "Passive Bci Based on Sustained Attention Detection: An fNIRS Study," in *Advances in Brain Inspired Cognitive Systems*. Editors C.-L. Liu, A. Hussain, B. Luo, K. C. Tan, Y. Zeng, and Z. Zhang (Springer International Publishing), 220–227. doi:10.1007/978-3-319-49685-6_20

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.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership