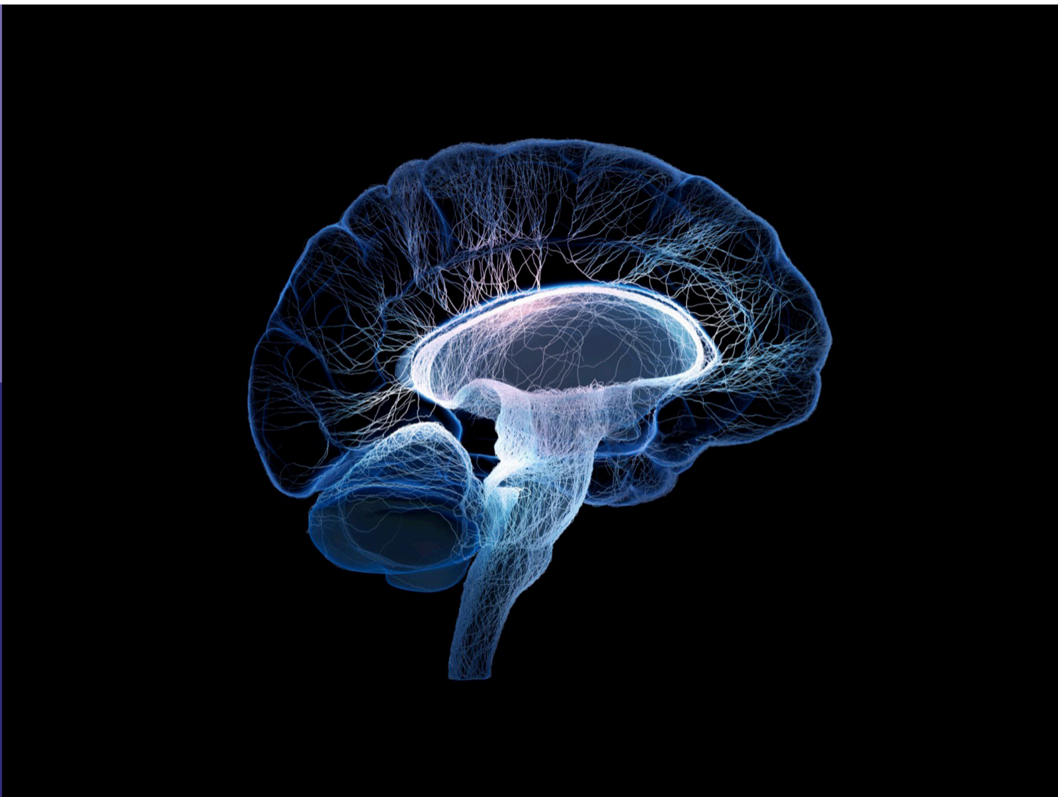
frontiers | Research Topics

Multimodal fusion technologies and applications in the context of neuroscience

Edited by
Xiaomao Fan, Yunpeng Cai, Weidong Gao, Raffaele Gravina,
Hui Zhou, Dingguo Zhang and Jian Huang

Published in
Frontiers in Neuroscience



ISSN 1664-8714
ISBN 978-2-8325-2586-9
DOI 10.3389/978-2-8325-2586-9

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

Multimodal fusion technologies and applications in the context of neuroscience

Topic editors
Xiaomao Fan — Shenzhen Technology University, China
Yunpeng Cai — Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences (CAS), China
Weidong Gao — Beijing University of Posts and Telecommunications (BUPT), China
Raffaele Gravina — University of Calabria, Italy
Hui Zhou — Nanjing University of Science and Technology, China
Dingguo Zhang — University of Bath, United Kingdom
Jian Huang — Huazhong University of Science and Technology, China

Citation
Fan, X., Cai, Y., Gao, W., Gravina, R., Zhou, H., Zhang, D., Huang, J., eds. (2023).
Multimodal fusion technologies and applications in the context of neuroscience.
Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-2586-9

Table of contents

05 Editorial: Multimodal fusion technologies and applications in the context of neuroscience
Xiaomao Fan, Meiyu Qiu, Weidong Gao, Wenjun Ma, Yunpeng Cai, Hui Zhou, Dingguo Zhang, Raffaele Gravina and Jian Huang

07 Upper Limbs Muscle Co-contraction Changes Correlated With the Impairment of the Corticospinal Tract in Stroke Survivors: Preliminary Evidence From Electromyography and Motor-Evoked Potential
Wenfei Sheng, Shijue Li, Jiangli Zhao, Yujia Wang, Zichong Luo, Wai Leung Ambrose Lo, Minghui Ding, Chuhuai Wang and Le Li

20 Flexible Non-contact Electrodes for Wearable Biosensors System on Electrocardiogram Monitoring in Motion
Xin Wang, Shuting Liu, Mingxing Zhu, Yuchao He, Zhilong Wei, Yingying Wang, Yangjie Xu, Hongguang Pan, Weimin Huang, Shixiong Chen and Guanglin Li

29 Towards Evaluating Pitch-Related Phonation Function in Speech Communication Using High-Density Surface Electromyography
Mingxing Zhu, Xin Wang, Hanjie Deng, Yuchao He, Haoshi Zhang, Zhenzhen Liu, Shixiong Chen, Mingjiang Wang and Guanglin Li

46 Multi-person feature fusion transfer learning-based convolutional neural network for SSVEP-based collaborative BCI
Penghai Li, Jianxian Su, Abdelkader Nasreddine Belkacem, Longlong Cheng and Chao Chen

57 Corrigendum: Multi-person feature fusion transfer learning-based convolutional neural network for SSVEP-based collaborative BCI
Penghai Li, Jianxian Su, Abdelkader Nasreddine Belkacem, Longlong Cheng and Chao Chen

58 A spatio-temporal learning-based model for sleep apnea detection using single-lead ECG signals
Junyang Chen, Mengqi Shen, Wenjun Ma and Weiping Zheng

72 Resting-state electroencephalography changes in poststroke patients with visuospatial neglect
Yichen Zhang, Linlin Ye, Lei Cao and Weiqun Song

85 Unsupervised layer-wise feature extraction algorithm for surface electromyography based on information theory
Mingqiang Li, Ziwen Liu, Siqi Tang, Jianjun Ge and Feng Zhang

96 MMASleepNet: A multimodal attention network based on electrophysiological signals for automatic sleep staging
Zheng Yubo, Luo Yingying, Zou Bing, Zhang Lin and Li Lei

107 **Neural substrates in patients with visual-spatial neglect recovering from right-hemispheric stroke**
Lei Cao, Linlin Ye, Huanxin Xie, Yichen Zhang and Weiqun Song

118 **Prediction of balance function for stroke based on EEG and fNIRS features during ankle dorsiflexion**
Jun Liang, Yanxin Song, Abdelkader Nasreddine Belkacem, Fengmin Li, Shizhong Liu, Xiaona Chen, Xinrui Wang, Yueyun Wang and Chunxiao Wan

127 **E2SGAN: EEG-to-SEEG translation with generative adversarial networks**
Mengqi Hu, Jin Chen, Shize Jiang, Wendi Ji, Shuhao Mei, Liang Chen and Xiaoling Wang

144 **Emotion recognition based on multi-modal physiological signals and transfer learning**
Zhongzheng Fu, Boning Zhang, Xinrun He, Yixuan Li, Haoyuan Wang and Jian Huang

159 **Usefulness of phase gradients of otoacoustic emissions in auditory health screening: An exploration with swept tones**
Xin Wang, Mingxing Zhu, Yuchao He, Zhenzhen Liu, Xin Huang, Hongguang Pan, Mingjiang Wang, Shixiong Chen, Yuan Tao and Guanglin Li

171 **A survey of data element perspective: Application of artificial intelligence in health big data**
Honglin Xiong, Hongmin Chen, Li Xu, Hong Liu, Lumin Fan, Qifeng Tang and Hsunfang Cho

181 **Ascertaining the optimal myoelectric signal recording duration for pattern recognition based prostheses control**
Mojisola Grace Asogbon, Oluwarotimi Williams Samuel, Ejay Nsugbe, Yongcheng Li, Frank Kulwa, Deogratias Mzurikwao, Shixiong Chen and Guanglin Li

frontiers | Frontiers in Neuroscience

TYPE Editorial
PUBLISHED 16 May 2023
DOI 10.3389/fnins.2023.1213207

Editorial: Multimodal fusion technologies and applications in the context of neuroscience

Xiaomao Fan[1]*, Meiyu Qiu[1], Weidong Gao[2]*, Wenjun Ma[3], Yunpeng Cai[4], Hui Zhou[5], Dingguo Zhang[6], Raffaele Gravina[7] and Jian Huang[8]

[1]College of Big Data and Internet, Shenzhen Technology University, Shenzhen, China, [2]School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China, [3]School of Computer Science, South China Normal University, Guangzhou, China, [4]Research Center for Biomedical Information Technology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, [5]School of Automation, Nanjing University of Science and Technology, Nanjing, China, [6]Center for Biosensors, Bioelectronics and Biodevices, University of Bath, Bath, United Kingdom, [7]Department of Informatics, Modeling, Electronics and Systems, University of Calabria, Rende, Italy, [8]School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China

OPEN ACCESS

EDITED AND REVIEWED BY
Michele Giugliano,
International School for Advanced Studies (SISSA), Italy

*CORRESPONDENCE
Xiaomao Fan
✉ fanxiaomao@sztu.edu.cn
Weidong Gao
✉ gaoweidong@bupt.edu.cn

RECEIVED 27 April 2023
ACCEPTED 03 May 2023
PUBLISHED 16 May 2023

CITATION
Fan X, Qiu M, Gao W, Ma W, Cai Y, Zhou H, Zhang D, Gravina R and Huang J (2023) Editorial: Multimodal fusion technologies and applications in the context of neuroscience. *Front. Neurosci.* 17:1213207. doi: 10.3389/fnins.2023.1213207

KEYWORDS

multimodal data fusion, biomedical signals, data mining and knowledge discovery, deep learning applications, neuroscience

Editorial on the Research Topic
Multimodal fusion technologies and applications in the context of neuroscience

In recent years, sensor and information technologies have greatly boosted the wearable/portal/medical devices development. A large number of multimodal biomedical signals such as electroencephalography (EEG), electrocardiography (ECG), electrooculogram (EOG), and electromyography (EMG), have been recorded for rehabilitation analysis, mental disorders evaluation, emotion recognition, cardiovascular disease diagnosis, etc. In these research fields, most researchers often use single-modal biomedical signals to build the corresponding analysis models. However, many clinical practice tasks, such as disease diagnosis, arrhythmias detection, and sleep condition monitoring, require multimodal biomedical signals together to make correct diagnoses, decisions, identifications, and predictions. It is noted that learning from multimodal biomedical signals can offer the possibility of capturing corresponded information and gaining an in-depth understanding of the relationship among different modalities.

The aim of this topic is to present recent research works to advance the fundamental theory and technologies in biomedical signal processing methods, multimodal fusion algorithms, and biomedical signal-based clinical applications. The special section began with several original researches about the applications of the biomedical signals in specified diseases such as stroke, dysphonia, and visual-spatial neglect (VSN). Three papers explored the applications of EMG signals. Sheng et al. explored the relevance between the increased muscle co-contraction and the corticospinal tract (CST) function in stroke survivors via EMG signals. It demonstrated that the CST and peripheral muscle co-contraction were closely related in stroke survivors. And increasing the intervention of the CST excitability would facilitate the recovery of muscle coordination in the upper limb after stroke. Zhu et al. have measured the speech and the high-density surface EMG signals of the subjects, which suggested that the muscle contraction patterns would be used as a reference of pitch-related phonation functions evaluation. It was a potential alternative method to improve the clinical method for evaluating muscle functions of dysphonia diagnoses, facial paralysis, and other neuromuscular-inclined diseases. Asogbon et al. used KNN, LDA, and RF algorithms to study the impact of various EMG-signal recording duration (SRD) on the characterization of motor intents related with multiple kinds of finger gestures. It demonstrated that choosing a optimal signal length was crucial to characterize multiple classes of targeted limb motions.

The applications of EEG signals in stroke have been explored in three papers. Liang et al. were the first who considered the EEG and functional near-infrared spectroscopy (fNIRS) features as the biomarkers for stroke assessment. The authors have established a linear regression model to predict Berg Balance Scale (BBS) values and used an eightfold cross validation to test the model. It got a result that the EEG features including stroke-related desynchronization (ERD), oxygenated hemoglobin (HBO), and the age were the promising biomarkers for stroke motor recovery. Two researches explored a most common cognitive impairment named visual-spatial neglect (VSN) of poststroke patients. Cao et al. explored the recovery neural substrates of VSN. The study had demonstrated that the dorsal attention networks played a more significant role in recovery from VSN instead of ventral attention networks and the cerebellum was also involved in recovery. Zhang et al. explored the resting-state EEG (rsEEG) features in stroke patients with VSN, which suggested that the resting-state DARAH could differentiate the patient with VSN or not. They demonstrated the resting-state EEG signals would be a useful tool for VSN stroke patients' monitoring and DAR and pdBSI alpha parameters in resting-state EEG could be useful biomarkers.

Aside from EEG and EMG signals, ECG signals are also critical indicators of disease detection. Existing measuring methods of ECG signals don't meet the demands of dynamic measurement. Wang et al. have developed a wearable biosensors system for dynamic ECG monitoring which used a flexible electrode. This system was able to collect high quality ECG signals when subjects exercised. It showed that the proposed electrode could be a potential tool used in long time detection for physiological signal measurements for patients and athletes.

To overcome the effect of physiological signals noises and the single signals' low accuracy, Fu et al. proposed a substructure-based joint probability domain adaptation algorithm (SSJPDA) with bi-projection matrix (BPM) algorithm. The authors used these algorithms to recognize the emotion of subjects based on multimodal fusion physiological data. Compared with other algorithms, the proposed SSJPDA and SSJPDA-BPM algorithms could better deal with noises in data and had improved the performance of emotion recognition.

Deep learning algorithms based on physiological signals were explored in four papers. Li et al. have constructed a centralized steady-state visually evoked potential collaborative brain computer interface (SSVEP-cBCI) system which studied the multi- person EEG features. The system used a transfer learning-based convolutional neural network and three feature fusion methods, which showed the multi-person fusion features achieve more competitive results than single person's. Hu et al. have proposed a generative adversarial network E2SGAN based on EEG-to-stereoelectroencephalography (SEEG), which was aim to synthesize SEEG data from the simultaneous EEG data. E2SGAN is superior to the baseline methods on the real-patient experiments, which demonstrated that the synthesized results had the potential to capture abnormal discharges of the epileptic patients before seizures.

The algorithms about diagnosing the sleep disorders were explored in the other two papers. Chen et al. have creatively proposed a novel method named CNN-BiGRU which consists of considerable spatio-temporal blocks. It was used to classify the sleep apnea (SA) events based on ECG signals. Compared with the state-of-art ECG-based detection methods, CNN-BiGRU demonstrated an obviously competitive result, which could provide sleep monitoring service for the SA detection. Yubo et al. have explored a multimodal attention network MMASleepNet for sleep staging, which can extract the effective features from the multimodal electrophysiological information. Compared with the baseline methods, MMASleepNet performs better in the accuracy and the training speed aspects. It provides a good solution for multimodal sleep monitoring.

Conclusion and further considerations

To sum up, the papers accepted by this Research Topic mainly explored the biomedical signals applications or novel algorithms based on the detection of diseases in neural science. It enriches the present research studies, some of papers proposed novel methods that achieved better results than baselines. By issuing this Research Topic, it greatly boosts the advancement of multimodal fusion technologies for neuroscience applications based on biomedical signals.

Author contributions

XF conceptualized this editorial and MQ wrote it. XF, WG, YC, and WM revised this editorial and others significantly contributed to it.

Funding

This work was partially supported by the Basic and Applied Research Project of Guangdong Province under Grant No. 2022B1515130009, the Special Subject on Agriculture and Social Development, Key Research and Development Plan in Guangzhou under Grant No. 2023B03J0172, and Key Projects of National Social Science Foundation of China under Grant No. 19ZDA041.

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

frontiers | Frontiers in Neuroscience

ORIGINAL RESEARCH
published: 01 June 2022
doi: 10.3389/fnins.2022.886909

Upper Limbs Muscle Co-contraction Changes Correlated With the Impairment of the Corticospinal Tract in Stroke Survivors: Preliminary Evidence From Electromyography and Motor-Evoked Potential

Wenfei Sheng[1†], Shijue Li[1†], Jiangli Zhao[1], Yujia Wang[2], Zichong Luo[2], Wai Leung Ambrose Lo[1], Minghui Ding[1], Chuhuai Wang[1] and Le Li[3*]*

[1] Department of Rehabilitation Medicine, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China, [2] Faculty of Science and Technology, University of Macau, Taipa, Macao SAR, China, [3] Institute of Medical Research, Northwestern Polytechnical University, Xi'an, China

Edited by:
Hui Zhou,
Nanjing University of Science and Technology, China

Reviewed by:
Xu Zhang,
University of Science and Technology of China, China
Yi Zhu,
Nanjing Medical University, China

***Correspondence:**
Chuhuai Wang
wangchuh@mail.sysu.edu.cn
Le Li
lile5@nwpu.edu.cn

† These authors have contributed equally to this work

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

Received: *01 March 2022*
Accepted: *25 April 2022*
Published: *01 June 2022*

Citation:
Sheng W, Li S, Zhao J, Wang Y, Luo Z, Lo WLA, Ding M, Wang C and Li L (2022) Upper Limbs Muscle Co-contraction Changes Correlated With the Impairment of the Corticospinal Tract in Stroke Survivors: Preliminary Evidence From Electromyography and Motor-Evoked Potential. Front. Neurosci. 16:886909. doi: 10.3389/fnins.2022.886909

Objective: Increased muscle co-contraction of the agonist and antagonist muscles during voluntary movement is commonly observed in the upper limbs of stroke survivors. Much remain to be understood about the underlying mechanism. The aim of the study is to investigate the correlation between increased muscle co-contraction and the function of the corticospinal tract (CST).

Methods: Nine stroke survivors and nine age-matched healthy individuals were recruited. All the participants were instructed to perform isometric maximal voluntary contraction (MVC) and horizontal task which consist of sponge grasp, horizontal transportation, and sponge release. We recorded electromyography (EMG) activities from four muscle groups during the MVC test and horizontal task in the upper limbs of stroke survivors. The muscle groups consist of extensor digitorum (ED), flexor digitorum (FD), triceps brachii (TRI), and biceps brachii (BIC). The root mean square (RMS) of EMG was applied to assess the muscle activation during horizontal task. We adopted a co-contraction index (CI) to evaluate the degree of muscle co-contraction. CST function was evaluated by the motor-evoked potential (MEP) parameters, including resting motor threshold, amplitude, latency, and central motor conduction time. We employed correlation analysis to probe the association between CI and MEP parameters.

Results: The RMS, CI, and MEP parameters on the affected side showed significant difference compared with the unaffected side of stroke survivors and the healthy group. The result of correlation analysis showed that CI was significantly correlated with MEP parameters in stroke survivors.

Conclusion: There existed increased muscle co-contraction and impairment in CST functionality on the affected side of stroke survivors. The increased muscle co-contraction was correlated with the impairment of the CST. Intervention that could improve the excitability of the CST may contribute to the recovery of muscle discoordination in the upper limbs of stroke survivors.

Keywords: stroke, muscle co-contraction, motor-evoked potential, corticospinal tract, correlation analyses

INTRODUCTION

Stroke is the major disease that leads to mortality and disability worldwide GBD 2016 Stroke Collaborators (2019). The most common impairment of stroke survivors is motor impairment, which affects an individual's ability to perform everyday activities and participate in social life (Langhorne et al., 2009). Hemiparesis is the most common symptom in stroke survivors (Bourbonnais et al., 1989; Nakayama et al., 1994; Wolfe, 2000; Roger et al., 2012), with abnormal muscle activation patterns being commonly observed (Bourbonnais et al., 1989). In many stroke survivors, motor impairment originates primarily in abnormal muscle coactivation (Dewald et al., 2001). Muscle co-contraction refers to the simultaneous activity of the agonist and antagonist muscles across the same joint (Banks et al., 2017; Souissi et al., 2018). Surface electromyography (EMG) can detect the muscle activities of the agonist and antagonist muscles (Campanini et al., 2020), and it can be used to identify abnormal muscular coordination in stroke survivors (Bourbonnais et al., 1989; Safavynia et al., 2011). The co-contraction between agonist and antagonist muscles can be evaluated quantitatively using the co-contraction index (CI) (Frost et al., 1997; Song and Tong, 2013; Banks et al., 2017; Li et al., 2020). Song and Tong (2013) found that there was an increased co-contraction between agonist and antagonist muscles of elbow during voluntary movement on the affected side compared with the unaffected side in stroke survivors. Increased muscle co-contraction indicates that the muscles could not contract independently (Hu et al., 2013). Hammond et al. (1988) found that the agonist and antagonist muscles of the wrist joint have a higher co-contraction ratio during voluntary isometric contraction on the affected side compared with the healthy control group. Kamper and Rymer (2001) found that stroke survivors had excessive co-contraction of hand muscles compared with the healthy control group. Increased muscle co-contraction leads to impairment in the upper limb motor function in stroke survivors. Previous studies reported that increased muscle co-contraction had a negative effect on voluntary movement (Chae et al., 2002; Chalard et al., 2019). It could bring about increased duration of the movement, muscle discoordination, and decreased range of movement (Arene and Hidler, 2009; Gross et al., 2015; Sarcher et al., 2015). Several studies applied the CI to evaluate muscular coactivation pattern changes during stroke recovery (Hammond et al., 1988; Chae et al., 2002; Hu et al., 2009; Nam et al., 2017; Qian et al., 2017; Rong et al., 2017). Chae et al. (2002) found that the co-contraction between the agonist and antagonist muscles of the wrist showed a negative relationship to motor function of the upper limbs, evaluated by Fugl-Meyer scales and arm motor ability test. Previous studies assessed the structural and functional muscle alternation after stroke by ultrasonography (Kim et al., 2021), muscle biopsy (Dalise et al., 2020), sEMG (Hu et al., 2015), high-density-surface (HD-sEMG) (Tanzarella et al., 2020), and dual-energy X-ray absorptiometry (Choi et al., 2021). There are studies that applied sEMG, kinematic parameters, and clinical scales to evaluate the upper-limb motor function in stroke survivors (Donoso Brown et al., 2014; Pan et al., 2021). But these studies focused only on the changes in the properties of muscles. For better stroke rehabilitation, it is necessary to assess the peripheral muscle changes and alternation in descending motor pathway at the same time (Azzollini et al., 2021).

The corticospinal tract (CST) is the principal neural pathway of the voluntary drive to the upper limb where muscle synergy is modulated (Lemon, 2008; McMorland et al., 2015; Van Wittenberghe and Peterson, 2021). The assessment of CST includes the transcranial magnetic stimulation (TMS) and diffusion tensor imaging (DTI) (Jang, 2013; Potter-Baker et al., 2016). Motor-evoked potential (MEP), elicited by TMS, provides quantitative method for evaluating the functional integrity of the CST (Groppa et al., 2012; Bestmann and Krakauer, 2015; Okamoto et al., 2021). TMS could induce rapidly changing magnetic field that stimulates cortical neurons and generates induced current. The induced current then depolarizes cortical axons and triggers MEP at suprathreshold stimulus intensities. The MEP is transmitted to the peripheral muscle through a descending path such as CST and corticobulbar motor pathways (Groppa et al., 2012). MEP provides insight into the mechanisms of motor output control (Bestmann and Krakauer, 2015) and can be applied to monitor the clinical progression stroke recovery (Cakar et al., 2016). Longer latency, smaller amplitude, and higher thresholds of MEP were observed on the affected side compared with the unaffected side in stroke survivors (Turton et al., 1996; Pennisi et al., 2002). During the recovery from stroke, the MEP of the paresis side changes toward the healthy state (Barker et al., 2012). Byrnes et al. (1999) showed that the MEP had a broad relationship with motor deficit as assessed by the Motor Assessment Scale and British Medical Research Council Scale (Brouwer and Schryburt-Brown, 2006). Bowden et al. (2014) showed that muscle weakness of the upper limb in stroke survivors resulted from the impairment of the descending corticospinal connections. Madhavan et al. (2011) investigated the correlation between the CST integrity and muscle strength by TMS, DTI, and dynamometer. There are few studies combining the assessment of muscle activation with the evaluation of the CST. Although there are many studies using MEP to assess the motor function of the stroke survivors (Turton et al., 1996; Traversa et al., 1998; Hendricks et al., 2003), only a limited number of studies investigated the correlation between MEP and muscle discoordination in stroke survivors.

Hortobágyi and Devita (2006) found that there was increased muscle co-contraction in the agonist and antagonist muscles in older adults. The age-associated change in the muscle co-contraction might result from the cortical component. The increased coactivation between the ankle and knee extensors in the paretic leg of stroke survivors was correlated with alterations in propriospinal pathways (Dyer et al., 2011). Chalard et al. (2020) conducted an EEG study, which found that an increased co-contraction was correlated with cortical movement-related beta oscillation alterations. Increased recurrent Renshaw inhibition is considered to be related to the increased co-contraction of the agonist and antagonist muscles (Katz and Pierrot-Deseilligny, 1982), Another physiological mechanism associated with increased muscle co-contraction of the agonist and antagonist includes the decrease in the Ia reciprocal inhibition, presynaptic inhibition, and Ib inhibition (Morita et al., 2006; Crone et al., 2007; Baude et al., 2019). The decrease in reciprocal inhibition was associated with the impairment of the CST (Crone et al., 2004). Much remains to be understood about the correlation between the impairment of the CST and increased muscle co-contraction of the upper limbs of stroke survivors. The impaired motor function is not only the result of the dysfunction of central motor control system but the result of the alternation in muscle activation (Azzollini et al., 2021). MEP evoked by TMS could reflect the function of the CST. The sEMG data provided the information of the peripheral muscle activity. The correlation between sEMG and MEP from TMS could lead to a better understanding about the mechanism of the abnormal muscle contraction pattern in stroke survivors. Especially, these findings provided insights into the mechanism of increased muscle co-contraction in stroke survivors. Therefore, the study aimed to probe the possible correlation between MEP and CI of the agonist and antagonist muscles during voluntary movement of the upper limbs of stroke survivors. We attempted to investigate whether the abnormal muscle coordination was associated with the impairment of the CST in stroke survivors.

MATERIALS AND METHODS

Participants

Nine stroke survivors and nine age-matched healthy people were recruited after obtaining approval from the Human Subjects Ethics Subcommittee of The First Affiliated Hospital of Sun Yat-sen University. This study is part of the clinical research that was registered on the Chinese Clinical Trial Registry (ChiCTR2000032245). All participants signed written consent prior to participation. The study was performed in accordance with the Declaration of Helsinki. The inclusion criteria of stroke survivors are (1) unilateral stroke; (2) 30–75 years old; (3) the elbow flexors, wrist flexors, and finger flexors scored less than 3 on the Modified Ashworth Scale (MAS); (4) the muscle strength of elbow extensors, wrist extensors, and finger extensors scored more than 2 on the manual muscle testing; (5) no metal implants in brain and cervical spine; (6) has sufficient cognitive ability to follow experimental procedure; and (7) has detectable MEP on abductor pollicis brevis. The exclusion criteria are (1) epilepsy, (2) pregnancy, (3) severe respiratory and circulatory failure, and (4) posterior circulation infarction or posterior circulation hemorrhage.

Clinical Measures

Fugl-Meyer Assessment Upper Extremity Scale (FMA-UE) and Action Research Arm Test (ARAT) were applied to assess motor function of the upper limb of stroke survivors. The MAS was applied to evaluate the spasticity. All survivors were evaluated by an experienced therapist.

Electromyography Experiment

After preparing the skin (Nuprep, Weaver and Company, Aurora, CO, United States) and 75% alcohol, four pairs of surface electrodes (Dongguan Quanding Medical Supplies Co., Ltd., Guangdong, China) were placed on the skin surface of four muscle groups in the upper limbs to record EMG signals. The involved muscle groups included flexor digitorum (FD), extensor digitorum (ED), the biceps brachii (BIC), and the triceps brachii (TRI) muscle groups (**Figure 1A**).

Participants were first directed to perform isometric maximal voluntary contraction (MVC) of the four involved muscles. When a participant conducted the MVC test for the ED and FD, the elbow was kept extended at 130° and the wrist was kept in a neutral position. When a participant conducted the MVC test for the BIC and TRI, the shoulder joint was kept flexed at 45° and the elbow flexed at 90° (Hu et al., 2013; Nam et al., 2017). Each contraction was repeated three times with 2 min intervals. EMG signals were recorded during each MVC test. The EMG data recorded in MVC were used to normalize the RMS recorded in horizontal task.

Then, all the participants were guided to conduct a horizontal task (**Figure 1A**) for left upper extremity, the participant was instructed to grasp a sponge at point A. Then, the participant held the sponge and transferred it laterally to point B with natural speed. Finally, the participant released the sponge at point B. For right upper extremity, the participant performed the sponge transfer from point B to point A. The distance between point A and point B is 50 cm. The thickness of the sponge is 3 cm. EMG signals were recorded during the whole horizontal task. During the transfer task, the participant kept the testing hand at a height of 2–5 cm from the table top. Each task was repeated three times with 2 min intervals. **Figure 1A** shows the setup of the horizontal task.

Motor-Evoked Potential Experiment

Figure 1B shows the setup of the MEP evaluation. A pair of HEX Dual Electrodes (Noraxon U.S.A. Inc., Arizona, United States) were attached to the surface of the abductor pollicis brevis to record the MEP signals. We employed a magnetic stimulator (Yiruide CCY-IA, Wuhan, China) to deliver magnetic stimulation. The stimulation coil was an "8"-shaped coil with a radius of 7 cm. The maximum stimulator output (MSO) of

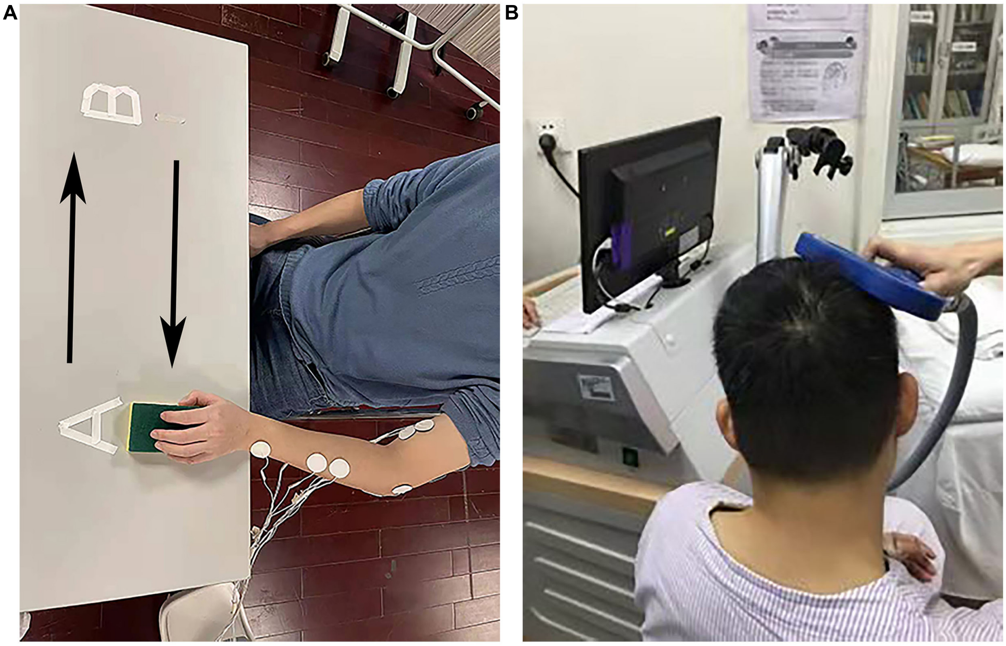


FIGURE 1 | The experimental setup. **(A)** The electromyography (EMG) evaluation during horizontal task. **(B)** The experimental setup for motor-evoked potentials (MEP) evaluation.

the magnetic stimulator was 2.0 T. The stimulator coil was held over the thumb area of the contralateral motor cortex and C7 cervical spine to elicit MEP signals. The resting motor threshold (rMT), latency and amplitude, and central motor conduction time (CMCT) were measured in bilateral abductor pollicis brevis in the upper limbs of all participants following the published guidelines (Rossini et al., 1994; Kobayashi and Pascual-Leone, 2003; Cakar et al., 2016).

The participants sat on the chair with palms face up on legs. The rMT was defined as the lowest stimulus intensity to evoke liminal MEP, which has an amplitude of at least 50 μV in 5 of 10 continuously trials in resting state. The rMT was expressed as % maximum stimulator output (% MSO) (Rossini and Rossi, 2007). The stimulation intensity on motor cortex was set at 120% of the rMT. The stimulation intensity on C7 cervical spine was set at 80% of the rMT. The CMCT refers to the difference in latency between MEP elicited by the cortical stimulation and MEP produced by spinal stimulation. The MEP signals were captured and amplified with a resolution of 200 μV, signals were then filtered with a bandpass of 10 Hz–2 kHz, and a noise eliminator of 50 Hz (Groppa et al., 2012).

Data Processing and Analysis

All raw EMG data were amplified 1,000 times (amplifier: INA 333, Texas Instruments Inc., Dallas, TX, United States) and sampled with 1,000 Hz for digitization with a data acquisition card (DAQ, 6218 NI DAQ card; National Instruments Corp., Austin, TX, United States). Then, the digitized EMG signals were transferred to the computer for storage. The signals were processed by removing bias, bandpass filtering (bandwidth range from 40 to 490 Hz), full-wave rectification. Lowpass filtering (30 Hz cutoff frequency with fourth-order zero-phrase Butterworth filter) was applied to have the linear EMG envelopes. A typical trial of linear envelopes of the EMG signals captured during horizontal task is shown in **Figure 2**. EMG data recorded during horizontal tasks were normalized to the maximum value emerging during the MVC test of each muscle. The maximum value of EMG activity was often found in the MVC test, but when higher amplitude of EMG activation emerged in horizontal task, this value would be substituted to be used for normalization. The muscle firing moment was identified as the abrupt moment of EMG signal activation that was greater than the threshold value (**Figure 2**). The threshold value is calculated by the mean of the EMG baseline plus two times the standard deviation, lasting for 20 ms.

The root mean square (RMS) was calculated to assess the EMG activation level of each muscle (Chae et al., 2002). The RMS value was computed from the EMG data section from the firing moment to the point when the task was finished. The RMS value was calculated by the following expression:

$$\mathrm{V_{RMS}} = \left[\frac{1}{T}\int v^2 \mathrm{dt}\right]^{1/2}$$

Where T is the length of the signal, and v is the voltage of the EMG signal (Chae et al., 2002). The RMS value could be applied to evaluate the neuromuscular system and the workload on muscles (Moritani, 1993; Shimomura et al., 1999). CI was calculated to assess the degree of muscle co-contraction (Hu et al., 2012, 2013; Nam et al., 2017). The CI value was

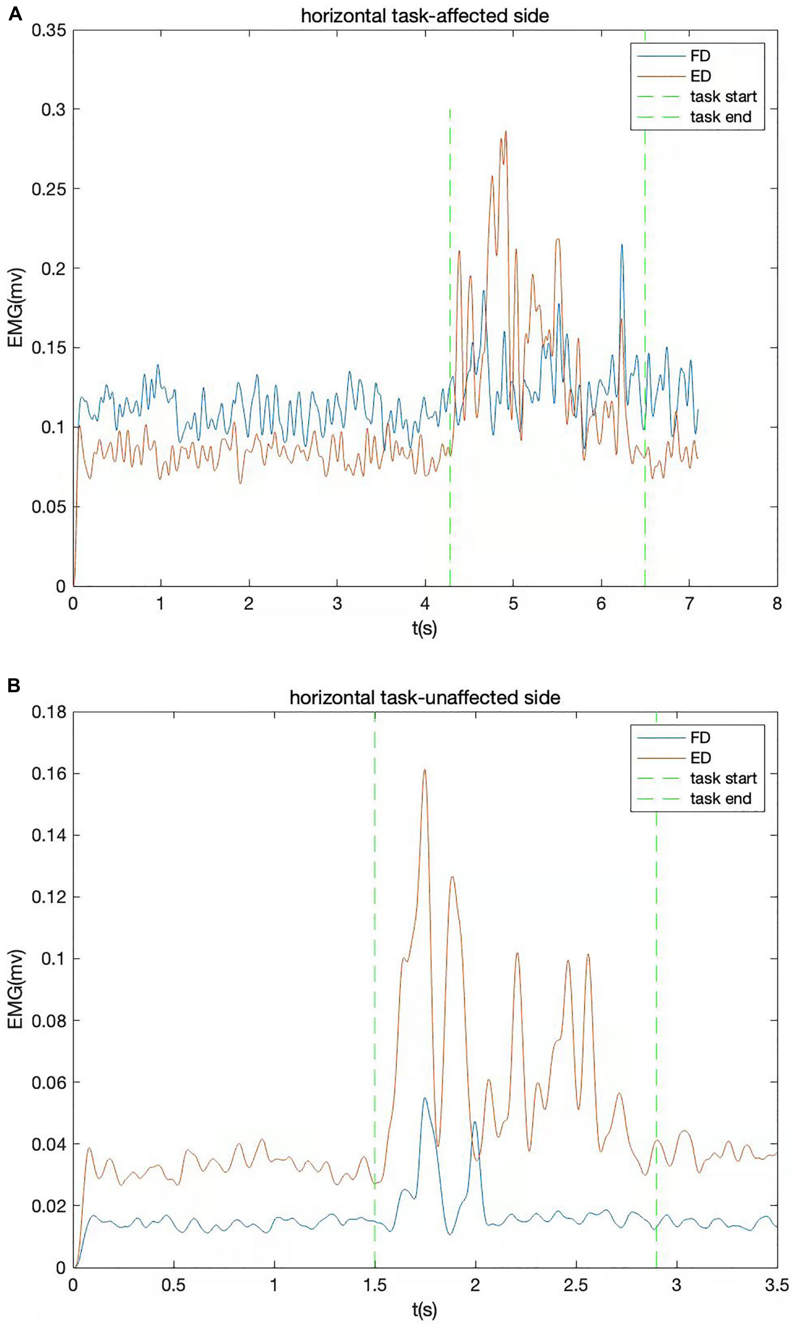


FIGURE 2 | The EMG signals of a typical trial from one stroke survivor captured during horizontal task. The EMG signal of flexor digitorum (FD) and extensor digitorum (ED) are shown together. **(A)** Affected side; **(B)** unaffected side.

calculated by the formula previously presented in Frost's study (Frost et al., 1997):

$$CI = \frac{1}{T}\int_{0}^{T} A_{ij}(t)dt$$

Where A_{ij} (t) is the overlapping activity of the muscles of i and j in the EMG envelopes, and T is the duration of the task. The CI value varied from 0 to 1. When the activities of two muscles were fully overlapping and the level of EMG activity kept at 1 during the task, the CI value reached 1. When the activities of two muscles did not overlap at all during the task, the CI value was 0.

Statistical Analysis

The Shapiro–Wilk test was employed to verify the data normality. The difference of EMG parameters and MEP parameters among the groups (affected side, unaffected side, and healthy group) was analyzed by one-way analysis of variance (ANOVA). Bonferroni test was applied to perform *post hoc* pairwise comparison. When the data did not conform to a normal distribution, we transformed the data by natural logarithm or taking cosine until the data were normally distributed. Correlation analysis was applied to analyze the correlation among clinical scale outcomes, CI, and MEP parameters. For data with a normal distribution, Pearson's correlation analysis was adopted. For data that do not conform to a normal distribution, Spearman's correlation analysis was employed. The significant level was set at $p < 0.05$. The data analysis was performed with the software SPSS Statistics 26 (IBM Inc., Seattle, WA, United States).

RESULTS

The characteristics of stroke survivors included in the study are summarized in **Table 1**.

Electromyography Parameters

Figure 3 presents the mean and standard deviation of normalized RMS values of the four muscles (FD, ED, BIC, and TRI) during horizontal task. The RMS values of the muscles (FD, BIC, and TRI) on the affected side were significantly higher than the unaffected side in stroke survivors ($p < 0.05$) and the healthy group ($p < 0.05$). The RMS values of the four muscles of the unaffected side of stroke survivors were higher than healthy people but without a significant difference ($p > 0.05$).

Table 2 presents the CI of ED and FD. The CI between two muscles of ED and FD during horizontal task on the affected side was significantly higher than those from the unaffected side ($p < 0.05$) in stroke survivors and the healthy group ($p < 0.05$). The CI of the two muscles of ED and FD in the unaffected side of survivors were higher than those in healthy group but do not reach significant level ($p > 0.05$).

Motor-Evoked Potential

Table 2 shows the mean and standard deviation of MEP parameters among the three groups. The MEP latency of the affected side of stroke survivors was significantly longer than the unaffected sides ($p < 0.05$) and the dominant side of healthy people ($p < 0.05$). The MEP amplitude of the affected side of stroke survivors was significantly lower than the unaffected sides ($p < 0.05$) and the dominant side of healthy people ($p < 0.05$). The MEP threshold of the affected side of stroke survivors was significantly higher than the unaffected sides ($p < 0.05$) and the dominant side of healthy people ($p < 0.05$). The MEP CMCT of the affected side of the survivors were significantly longer than the unaffected sides ($p < 0.05$) and the dominant side of healthy people ($p < 0.05$). There was no significant difference in all four MEP parameters between the unaffected sides with the dominant side of healthy people ($p > 0.05$).

Motor-Evoked Potential and Clinical Scales

Table 3 presents the correlation between MEP parameters and the outcomes of clinical scales. The latency and CMCT of MEP on the affected side were negatively correlated with the upper-limb motor function that was assessed by both clinical scales of Fugl-Meyer scale and ARAT. The amplitude and rMT of MEP showed no significant correlation with the outcome of both clinical scales of Fugl-Meyer scale and ARAT (**Figure 4** and **Table 3**).

Root Mean Square and Motor-Evoked Potential

The latency of MEP was positively correlated with the RMS value of EMG activity of FD ($r = 0.52$, $p = 0.033$), BIC ($r = 0.667$, $p = 0.003$), and TRI ($r = 0.578$, $p = 0.015$) muscles during horizontal task in stroke survivors. The RMS value of ED during horizontal task was not significantly correlated with the latency of MEP ($r = 0.245$, $p = 0.343$).

Co-contraction Index and Motor-Evoked Potential

Figure 5 presents the correlation between CI and MEP parameters. The CI of the two muscles of ED and FD during horizontal task was positively correlated with latency ($r = 0.7$, $p < 0.05$), rMT ($r = 0.52$, $p < 0.05$), and CMCT ($r = 0.56$, $p < 0.05$) of the MEP in stroke survivors. The CI of the muscles of ED and FD increased along with the increase in latency, rMT, and CMCT of the MEP. The CI of the two muscles of ED and FD during horizontal task was negatively correlated with the amplitude of the MEP ($r = -0.55$, $p < 0.05$) in stroke survivors.

TABLE 1 | Clinical characteristics of stroke survivors.

Subject	Age	Gender	Paralyzed side	Stroke type	Month since stroke	FMA-UE	ARAT	MAS (wrist)
1	57	Male	Left	Ischemia	4	53	38	1
2	68	Male	Left	Hemorrhage	2	66	57	0
3	38	Male	left	Hemorrhage	9	32	14	1+
4	31	Male	Left	Hemorrhage	2	61	35	0
5	69	Male	Right	Ischemia	2	54	38	1+
6	72	Female	Left	Ischemia	2	49	33	1
7	66	Male	Left	Ischemia	2	63	54	0
8	31	Male	Left	Hemorrhage	2	62	38	0
9	71	Male	Left	Ischemia	10	45	21	1+

FMA-UE, Fugl-Meyer Assessment Upper Extremity Scale; ARAT, Action Research Arm Test; MAS, Modified Ashworth Scale.

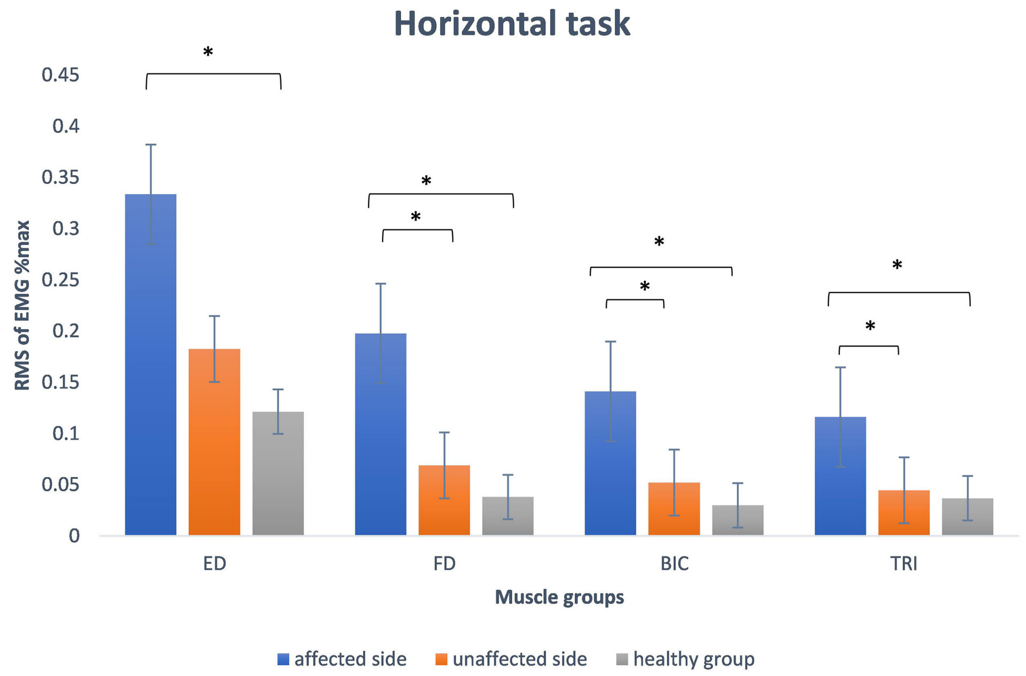


FIGURE 3 | The mean and standard deviation of root mean square of EMG data of all participants during horizontal task involved in the study. The significant difference between the groups is indicated by (*$p < 0.05$ with one-way ANOVA). ED, extensor digitorum; FD, flexor digitorum; BIC, biceps brachii; TRI, the triceps brachii.

The CI of the muscles of ED and FD increased as the amplitude of the MEP decreased.

DISCUSSION

The main finding of this study indicated that muscle activation alternation of the upper limb was correlated with the changes in the function of CST in stroke survivors. The result provided insights into the origin of the increased muscle co-contraction in the upper limb of stroke survivors. The latency and CMCT of MEP on the affected side in stroke survivors were negatively correlated with the Fugl-Meyer upper extremity scale and ARAT. The CI of the muscle pair of ED and FD during horizontal task was significantly correlated with MEP parameters in stroke survivors. The CI increased, along with an increase in latency, rMT, and CMCT of the MEP. The CI increased along with a reduction in the amplitude of the MEP in the upper limbs of stroke survivors.

The Differences of Root Mean Square Value and Co-contraction Index Among Groups

The RMS values recorded from the four muscles of the affected side were higher than the unaffected side in the stroke group and also the healthy group during horizontal task. Canning et al. (2000) found that there was a higher biceps muscle activation during a tracking task in the paretic upper limb of stroke survivors compared to the healthy group. Lee et al. (2015) reported an increase in muscle activations on the affected side during a drinking task compared with the unaffected side. Wagner et al. (2007) found that there was higher level of muscle activity during a reaching task on the affected side of the upper limb of stroke survivors than healthy individuals. The results of our study are consistent with previous studies that showed increased muscle activation on the affected side during task execution in stroke survivors. Hu et al. (2013) found a reduction in the upper limb muscles EMG activity that corresponded with upper limbs functional improvement after a 20 sessions training program. The reduction in EMG activities post training was proposed to be the result of a reduction in spasticity that contributed to a lower level of muscle activities (Hu et al., 2007, 2013). The other possible reason was that training sessions contributed to an increase in muscle force production which enabled survivors to perform tasks with less muscle effort (Hu et al., 2007). These findings supported that muscle activation level during task could assist the assessment of clinical progression in motor function during the recovery from stroke. The CI of the muscle pairs of FD and ED on the affected side was higher than the unaffected side in the stroke group and the healthy group during horizontal task. The result was in agreement with previous studies which recorded EMG signals during isometric movement (Hammond et al., 1988; Kamper and Rymer, 2001; Chae et al., 2002). Silva et al. (2014) found that the co-contraction ratio of the proximal upper extremity muscles increased during reaching movement on the affected side of stroke survivors. In contrast to the previous studies, we investigated the abnormal muscular coordination pattern during horizontal task which was more similar to the movement pattern of some daily activities. Song and Tong (2013) reported an increase in co-contraction between the BIC and TRI of the affected side during a tracking task. The excessive co-contraction between the agonist and antagonist reflects the reduction in the control of muscle activity of the affected side in stroke survivors (Song and Tong, 2013). Hu et al. (2013) showed that the CI of ED and FD during horizontal task decreased, along with an improvement in motor function during the recovery from stroke. The result indicated the CI of ED and FD during task could assist in the monitoring of clinical progression of the upper limbs during the recovery from stroke.

The Difference in Motor-Evoked Potential Parameters Among Groups

The latency, rMT, and CMCT of MEP of the affected side of the stroke group were higher than the unaffected side and the healthy group. The amplitude of MEP was smaller on the affected side. These results were consistent with previously published studies (Turton et al., 1996; Traversa et al., 1997, 1998; Byrnes et al., 1999; Pennisi et al., 2002; Brouwer and Schryburt-Brown, 2006; Barker et al., 2012). MEPs are efficacious in evaluating the CST functionality (Di Lazzaro et al., 1999). The difference of MEPs between groups indicated impairment in the CST functionality. Cakar et al. (2016) showed that latency was negatively correlated with functional outcomes improvement. Cakar et al.'s (2016) results were consistent with our study, where the latency of the MEP was negatively correlated with motor function of the upper limb of stroke survivors. Besides, Cakar et al. (2016) reported that the rMT and amplitude of MEP were correlated with the clinical outcomes of Brunnstrom motor stage, Motricity index, finger tapping test, and motor activity log. However, this study did not observe significant correlation between rMT and Fugl-Meyer scale and ARAT, or between MEP amplitude and Fugl-Meyer scale and ARAT. The amplitude of MEP could be influenced by various factors such as the intensity of stimulation, the condition of intent muscle, and the condition of EEG phase and power fluctuations (Rossini et al., 2015). The rMT of MEP could also be impacted by drugs, age, the intent muscle, and sleep-wake cycles (Groppa et al., 2012). These might be the potential explanations for the different results observed between Cakar et al.'s study and this study. The amplitude and rMT have relatively higher intraindividual variations.

TABLE 2 | The mean (M) and standard deviation (SD) of CI and MEP parameters of stroke survivors and the healthy group.

	Affected side (M ± SD)	Unaffected side (M ± SD)	Healthy group (M ± SD)
CI (horizontal task)	0.69 ± 0.13[ab]	0.46 ± 0.18	0.44 ± 0.11
MEP latency (ms)	27.59 ± 5.14[ab]	22.15 ± 1.67	22.19 ± 2.23
Amplitude (μV)	196.02 ± 163.68[ab]	565.48 ± 334.98	366.85 ± 119.71
rMT (% MSO)	70.11 ± 19.64[ab]	39.78 ± 6.67	47.33 ± 8.31
CMCT (ms)	12.46 ± 6.12[ab]	7.32 ± 2.20	7.81 ± 1.33

MEP, motor-evoked potential; rMT, resting motor threshold; CMCT, central motor conduction time.
[a] *$p < 0.05$ compared to the unaffected side.*
[b] *$p < 0.05$ compared to the dominant side of healthy group.*

TABLE 3 | Correlation analyses between the motor function measures of affected limb, CI, and the MEP parameters of affected hemisphere.

	FMA-UE	ARAT
CI	−0.318 (*P* = 0.405)	−0.433 (*P* = 0.245)
MEP latency (ms)	−0.883 (*P* = 0.02)	−0.729 (*P* = 0.026)
Amplitude (μV)	0.117 (*P* = 0.765)	0.153 (*P* = 0.695)
rMT (% MSO)	−0.406 (*P* = 0.279)	−0.504 (*P* = 0.166)
CMCT (ms)	−0.883 (*P* = 0.02)	−0.678 (*P* = 0.045)

FMA-UE, Fugl-Meyer Assessment Upper Extremity Scale; ARAT, Action Research Arm Test; CI, co-contraction index; MEP, motor-evoked potential; rMT, resting motor threshold; CMCT, central motor conduction time.

Okamoto et al. (2021) found that the CMCT had a negative relationship with the ARAT score. Cakar et al. (2016) showed that the CMCT of MEP was correlated with the clinical outcomes of Barthel Index, Brunnstrom motor stage, finger tapping test, and motor activity log. Higher motor thresholds and smaller amplitude might result from a loss of corticomotoneurons in the corticospinal pathway and reduced excitability of the motor cortex (Byrnes et al., 1999; Brouwer and Schryburt-Brown, 2006). Thus, MEP latency and CMCT may be appropriate indicators to evaluate the motor function of the upper limb.

The Correlation Between Electromyography and Motor-Evoked Potential Parameters

Wagner et al. (2007) suggested that the increased level of muscle activities on the affected side of stroke survivors might have originated from extra muscle units recruitment owing to the loss of the functional motor units. In this study, the RMS values of the FD, BIC, and TRI during horizontal task were correlated with the latency of MEP. The results indicated that the increased muscle activity during voluntary movement was correlated with the impairment of the CST. The RMS value of ED showed no significant correlation with the latency of MEP. The possible reason was that the ED muscle took part in a small part of the horizontal task.

The latency of MEP is thought to reflect the conduction time for the neural impulses from the cortex to peripheral muscles (Groppa et al., 2012; Bestmann and Krakauer, 2015). In this study, the CI of FD and ED during horizontal task were positively correlated with the latency of MEP. Shorter latency of MEP corresponded to smaller CI value, which reflected better muscle coordination. The MEP signals involved in the study were recorded from the muscle of the abductor pollicis brevis. Thus, the latency of MEP recorded at FD and ED would be also prolonged due to the anatomical location. The result suggested that the level of the muscular coordination in the upper limb of stroke survivors was correlated with the impairment of cortical transmission.

The result indicated muscle co-contraction of the upper limb of stroke survivors was correlated with the loss of corticospinal projections. The amplitude of MEP is considered to reflect the integrity of the CST and the excitability of the motor

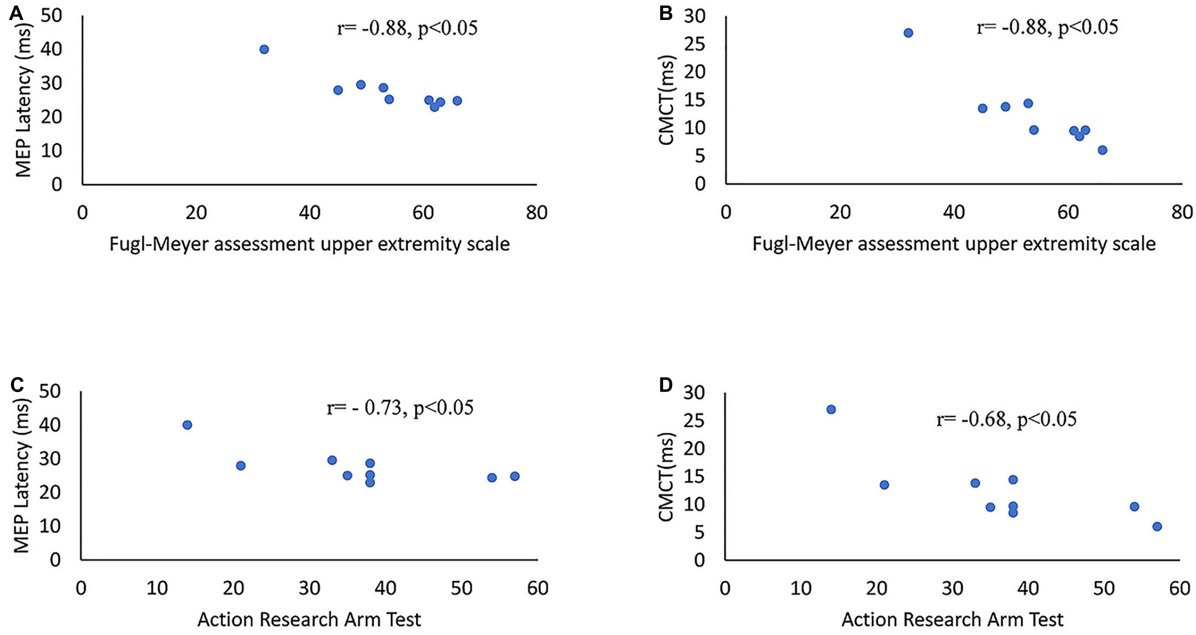


FIGURE 4 | Scatter diagrams of clinical scales and MEP parameters on the affected side of stroke survivors. **(A)** MEP latency versus Fugl-Meyer assessment upper extremity scales. **(B)** CMCT of MEP versus Fugl-Meyer assessment upper extremity scales. **(C)** MEP latency versus ARAT. **(D)** CMCT of MEP versus ARAT. ARAT, Action Research Arm Test; CMCT, central motor conduction time.

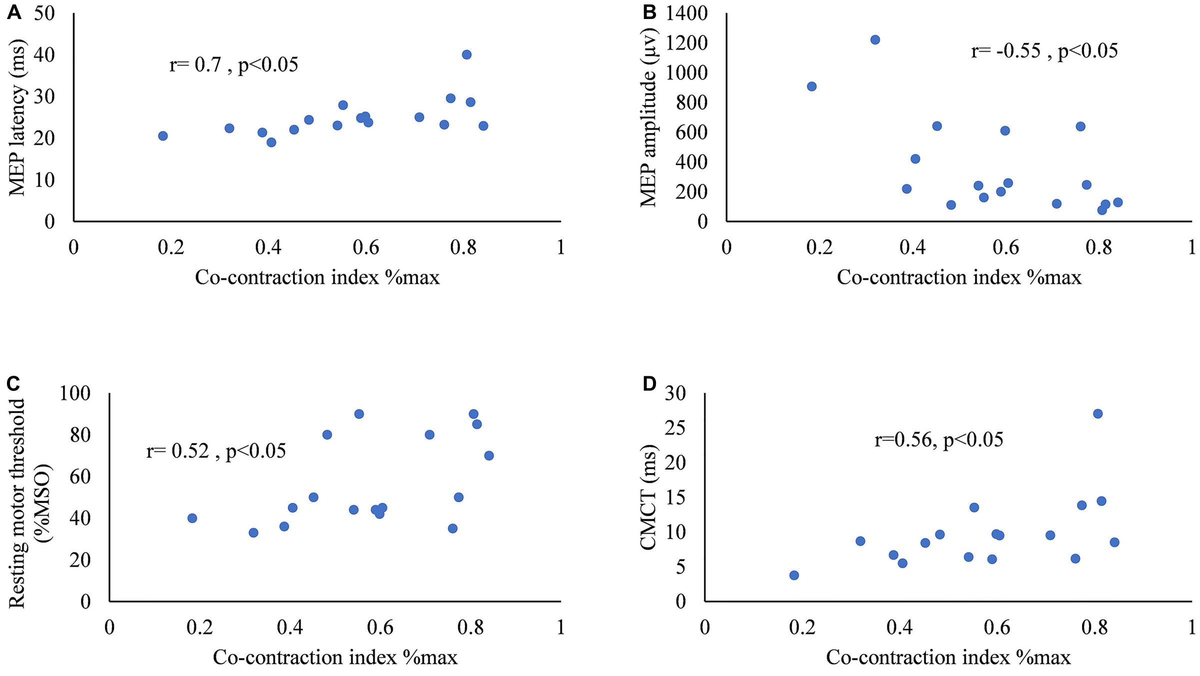


FIGURE 5 | Scatter diagrams of the co-contraction index of flexor digitorum and extensor digitorum during horizontal task versus **(A)** MEP latency, **(B)** amplitude of MEP, **(C)** resting motor threshold of MEP, and **(D)** central motor conduction time of MEP.

cortex (Groppa et al., 2012; Bestmann and Krakauer, 2015). The amplitude of MEP could reflect the transsynaptic excitation of corticospinal cells (Ziemann et al., 2015). Cakar et al. (2016) found that MEP amplitude was positively correlated with the outcomes of motor performance and dexterity of the upper limb. In our study, the CI of FD and ED during horizontal task was negatively correlated with the amplitude of MEP. The lower amplitude of MEP corresponded to an increase in CI. The reduction in MEP amplitude reflected the loss of corticospinal projections (Hömberg et al., 1991; Pennisi et al., 2002).

The rMT of MEP is considered to reflect the integrated excitability of the corticomotor projection (Groppa et al., 2012). Lower rMT was correlated with better motor performance (Cakar et al., 2016). For the CI of a muscle pair of a joint, lower CI value suggested a separation of the co-contraction phase which means the muscle pair could contract more independently (Hu et al., 2013). The CI of FD and ED during horizontal task was correlated with the rMT of MEP in our study. The increase in rMT could be caused by the decrease in the CST excitability which might have originated from loss of the corticospinal projection (Caramia et al., 1991; Pennisi et al., 2002). The result indicated the reduction in the CST excitability might be the possible mechanism to increase muscle co-contraction in the upper limbs of stroke survivors. Recently, Hammerbeck et al. (2021) found that the degree of the CST connectivity is the principal determinant of proximal dexterity strength and muscle synergy in upper limbs of patients with subacute stroke. The result of this study indicated that muscle co-contraction of FD and ED was correlated with the excitability of the CST.

Moreover, the result of this study provided evidence to support that CMCT positively correlated with muscle coherence. Longer CMCT corresponded to higher CI. The CMCT is the most related electrophysiological maker to evaluate the integrity of the CST (Groppa et al., 2012). The increase in CMCT might have originated from the loss of the fast corticospinal fibers and damage to the axonal (Misra and Kalita, 1995; Pennisi et al., 2002; Groppa et al., 2012). The result suggested that the increased muscle co-contraction of the upper limb was correlated with the loss of the corticospinal projection.

The correlation between CI and MEP parameters indicated that increased muscle co-contraction was correlated with the impairment of the CST in the upper limb of stroke survivors. Chae et al. (2002) suggested the underlying mechanisms for the increase in muscle co-contraction included an increase in alpha motoneuron excitability and increased activity in brainstem pathways after damage to the CST and cortical reorganization (Ohn et al., 2013). Chalard et al. (2020) found the increased muscle co-contraction was associated with a reduction in movement-related beta desynchronization. Other various physiological mechanisms such as reduction in Ia reciprocal inhibition, decrease in presynaptic inhibition, and reduction in Ib inhibition were considered to be correlated with an increase in muscle co-contraction in stroke survivors (Baude et al., 2019). The decrease in reciprocal inhibition was correlated with the impairment of the CST (Crone et al., 2004). We propose that the increased muscle co-contraction might have a cortical origin that was correlated with the impairment of the CST. Interventions that facilitate the recovery of the CST function might accelerate the recovery of the muscle coordination in the upper limb after stroke. An example of this type of intervention is neuromuscular electrical stimulation that could enhance the excitability of the CST and facilitate function recovery in stroke survivors (Ridding et al., 2001; Mang et al., 2010; Rong et al., 2015). Studies showed that neuromuscular electrical stimulation could improve the muscular coordination and clinical outcomes (Hu et al., 2009, 2012, 2013; Rong et al., 2015, 2017; Nam et al., 2017). The result of this study provided further evidence for the application of the neuromuscular electrical stimulation in the rehabilitation of muscle dyscoordination in the upper limb of stroke survivors. The correlation between CI and MEP parameters indicated that MEP could be applied to assess the muscle co-contraction in the upper limbs of stroke survivors.

Limitation

This is a preliminary study and the sample size is limited, which might contain type II error. Part of the survivors involved in the study were chronic stroke survivors. Further studies are required to verify the result in early stroke survivors. Stroke survivors at different stages could have different muscle contraction patterns. The stroke survivors at later stages such as more than 6 months could accompany with compensatory movement. The compensatory movement was proved to have originated from cortical compensatory neuroplasticity. The cortical compensatory neuroplasticity might influence the conclusion. The survivors involved in the study had mild to moderate impairment of the upper-limb function. Further studies are required to verify the results in stroke survivors with severe impairment. The stroke survivors with posterior circulation infarction or hemorrhage were excluded from the study. Future studies are required to verify the result in this group of stroke survivors.

CONCLUSION

We demonstrated a statistically significant correlation between muscle co-contraction and the CST function in stroke survivors. The correlation between CI and MEP parameters indicated the CST and peripheral muscle coordination were closely correlated in stroke survivors. Interventions that could increase the excitability of the CST might facilitate the recovery of muscle coordination in the upper limb after stroke.

DATA AVAILABILITY STATEMENT

The datasets analyzed during the current study are available from the corresponding authors upon reasonable request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Human Subjects Ethics Subcommittee of The First Affiliated Hospital of Sun Yat-sen University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LL and CW conceived and designed the study. WS, SL, and JZ performed the experiments. WS, YW, and ZL analyzed the data. WS, SL, and WL wrote the manuscript. MD, CW, and LL contributed to experiments. WL, MD, CW, and LL reviewed and edited the manuscript. All authors read and approved the manuscript.

FUNDING

This study was supported by the National Natural Science Foundation of China (Nos. 32071316, 82172532, and 31771016) and partly supported by the Guangdong Basic and Applied Basic Research Foundation (No. 2020A1515011356), the Guangzhou Science and Technology Project (Nos. 201803010083 and 201907010034), Fundamental Research Funds for the Central Universities (G2021KY05101), and Key Research and Development Project of Shaanxi Province (2022SF-117).

ACKNOWLEDGMENTS

We thank all the participants of this study.

REFERENCES

Arene, N., and Hidler, J. (2009). Understanding motor impairment in the paretic lower limb after a stroke: a review of the literature. *Top. Stroke Rehabil.* 16, 346–356. doi: 10.1310/tsr1605-346

Azzollini, V., Dalise, S., and Chisari, C. (2021). How does stroke affect skeletal muscle? state of the art and rehabilitation perspective. *Front. Neurol.* 12:797559. doi: 10.3389/fneur.2021.797559

Banks, C. L., Huang, H. J., Little, V. L., and Patten, C. (2017). Electromyography exposes heterogeneity in muscle co-contraction following stroke. *Front. Neurol.* 8:699. doi: 10.3389/fneur.2017.00699

Barker, R. N., Brauer, S. G., Barry, B. K., Gill, T. J., and Carson, R. G. (2012). Training-induced modifications of corticospinal reactivity in severely affected stroke survivors. *Exp. Brain Res.* 221, 211–221. doi: 10.1007/s00221-012-3163-z

Baude, M., Nielsen, J. B., and Gracies, J. M. (2019). The neurophysiology of deforming spastic paresis: a revised taxonomy. *Ann. Phys. Rehabil. Med.* 62, 426–430. doi: 10.1016/j.rehab.2018.10.004

Bestmann, S., and Krakauer, J. W. (2015). The uses and interpretations of the motor-evoked potential for understanding behaviour. *Exp. Brain Res.* 233, 679–689. doi: 10.1007/s00221-014-4183-7

Bourbonnais, D., Vanden Noven, S., Carey, K. M., and Rymer, W. Z. (1989). Abnormal spatial patterns of elbow muscle activation in hemiparetic human subjects. *Brain* 112 (Pt 1), 85–102. doi: 10.1093/brain/112.1.85

Bowden, J. L., Taylor, J. L., and Mcnulty, P. A. (2014). Voluntary activation is reduced in both the more- and less-affected upper limbs after unilateral stroke. *Front. Neurol.* 5:239. doi: 10.3389/fneur.2014.00239

Brouwer, B. J., and Schryburt-Brown, K. (2006). Hand function and motor cortical output poststroke: are they related? *Arch. Phys. Med. Rehabil.* 87, 627–634. doi: 10.1016/j.apmr.2006.02.006

Byrnes, M. L., Thickbroom, G. W., Phillips, B. A., Wilson, S. A., and Mastaglia, F. L. (1999). Physiological studies of the corticomotor projection to the hand after subcortical stroke. *Clin. Neurophysiol.* 110, 487–498. doi: 10.1016/s1388-2457(98)00044-3

Cakar, E., Akyuz, G., Durmus, O., Bayman, L., Yagci, I., Karadag-Saygi, E., et al. (2016). The relationships of motor-evoked potentials to hand dexterity, motor function, and spasticity in chronic stroke patients: a transcranial magnetic stimulation study. *Acta Neurol. Belg.* 116, 481–487. doi: 10.1007/s13760-016-0633-2

Campanini, I., Disselhorst-Klug, C., Rymer, W. Z., and Merletti, R. (2020). Surface emg in clinical assessment and neurorehabilitation: barriers limiting its use. *Front. Neurol.* 11:934. doi: 10.3389/fneur.2020.00934

Canning, C. G., Ada, L., and O'Dwyer, N. J. (2000). Abnormal muscle activation characteristics associated with loss of dexterity after stroke. *J. Neurol. Sci.* 176, 45–56. doi: 10.1016/s0022-510x(00)00305-1

Caramia, M. D., Cicinelli, P., Paradiso, C., Mariorenzi, R., Zarola, F., Bernardi, G., et al. (1991). 'Excitability changes of muscular responses to magnetic brain stimulation in patients with central motor disorders. *Electroencephalogr. Clin. Neurophysiol.* 81, 243–250. doi: 10.1016/0168-5597(91)90009-m

Chae, J., Yang, G., Park, B. K., and Labatia, I. (2002). Muscle weakness and cocontraction in upper limb hemiparesis: relationship to motor impairment and physical disability. *Neurorehabil. Neural Repair.* 16, 241–248. doi: 10.1177/154596830201600303

Chalard, A., Amarantini, D., Tisseyre, J., Marque, P., and Gasq, D. (2020). Spastic co-contraction is directly associated with altered cortical beta oscillations after stroke. *Clin. Neurophysiol.* 131, 1345–1353. doi: 10.1016/j.clinph.2020.02.023

Chalard, A., Amarantini, D., Tisseyre, J., Marque, P., Tallet, J., and Gasq, D. (2019). Spastic co-contraction, rather that spasticity, is associated with impaired active function in adults with acquired brain injury: a pilot study. *J. Rehabil. Med.* 51, 307–311. doi: 10.2340/16501977-2528

Choi, Y., Im, S., and Park, G. Y. (2021). Ultrasound evaluation of the rectus femoris for sarcopenia in patients with early subacute stroke. *J. Clin. Med.* 10:3010. doi: 10.3390/jcm10143010

Crone, C., Petersen, N. T., Gimenéz-Roldán, S., Lungholt, B., Nyborg, K., and Nielsen, J. B. (2007). Reduced reciprocal inhibition is seen only in spastic limbs in patients with neurolathyrism. *Exp. Brain Res.* 181, 193–197. doi: 10.1007/s00221-007-0993-1

Crone, C., Petersen, N. T., Nielsen, J. E., Hansen, N. L., and Nielsen, J. B. (2004). Reciprocal inhibition and corticospinal transmission in the arm and leg in patients with autosomal dominant pure spastic paraparesis (ADPSP). *Brain* 127, 2693–2702. doi: 10.1093/brain/awh319

Dalise, S., Azzollini, V., and Chisari, C. (2020). Brain and muscle: how central nervous system disorders can modify the skeletal muscle. *Diagnostics (Basel)* 10:1047. doi: 10.3390/diagnostics10121047

Dewald, J. P., Sheshadri, V., Dawson, M. L., and Beer, R. F. (2001). Upper-limb discoordination in hemiparetic stroke: implications for neurorehabilitation. *Top Stroke Rehabil.* 8, 1–12. doi: 10.1310/WA7K-NGDF-NHKK-JAGD

Di Lazzaro, V., Oliviero, A., Profice, P., Ferrara, L., Saturno, E., Pilato, F., et al. (1999). The diagnostic value of motor evoked potentials. *Clin. Neurophysiol.* 110, 1297–1307. doi: 10.1016/s1388-2457(99)00060-7

Donoso Brown, E. V., Mccoy, S. W., Fechko, A. S., Price, R., Gilbertson, T., and Moritz, C. T. (2014). Preliminary investigation of an electromyography-controlled video game as a home program for persons in the chronic phase of stroke recovery. *Arch. Phys. Med. Rehabil.* 95, 1461–1469. doi: 10.1016/j.apmr.2014.02.025

Dyer, J. O., Maupas, E., Melo Sde, A., Bourbonnais, D., and Forget, R. (2011). Abnormal coactivation of knee and ankle extensors is related to changes in heteronymous spinal pathways after stroke. *J. Neuroeng. Rehabil.* 8:41. doi: 10.1186/1743-0003-8-41

Frost, G., Dowling, J., Dyson, K., and Bar-Or, O. (1997). Cocontraction in three age groups of children during treadmill locomotion. *J. Electromyogr. Kinesiol.* 7, 179–186. doi: 10.1016/s1050-6411(97)84626-3

GBD 2016 Stroke Collaborators (2019). Global, regional, and national burden of stroke, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* 18, 439–458. doi: 10.1016/S1474-4422(19)30034-1

Groppa, S., Oliviero, A., Eisen, A., Quartarone, A., Cohen, L. G., Mall, V., et al. (2012). A practical guide to diagnostic transcranial magnetic stimulation: report of an IFCN committee. *Clin. Neurophysiol.* 123, 858–882. doi: 10.1016/j.clinph.2012.01.010

Gross, R., Leboeuf, F., Hardouin, J. B., Perrouin-Verbe, B., Brochard, S., and Rémy-Néris, O. (2015). Does muscle coactivation influence joint excursions during gait in children with and without hemiplegic cerebral palsy? Relationship between muscle coactivation and joint kinematics. *Clin. Biomech. (Bristol, Avon)* 30, 1088–1093. doi: 10.1016/j.clinbiomech.2015.09.001

Hammerbeck, U., Tyson, S. F., Samraj, P., Hollands, K., Krakauer, J. W., and Rothwell, J. (2021). The strength of the corticospinal tract not the reticulospinal tract determines upper-limb impairment level and capacity for skill-acquisition in the sub-acute post-stroke period. *Neurorehabil. Neural Repair.* 35, 812–822. doi: 10.1177/15459683211028243

Hammond, M. C., Fitts, S. S., Kraft, G. H., Nutter, P. B., Trotter, M. J., and Robinson, L. M. (1988). Co-contraction in the hemiparetic forearm: quantitative EMG evaluation. *Arch. Phys. Med. Rehabil.* 69, 348–351.

Hendricks, H. T., Pasman, J. W., Merx, J. L., Van Limbeek, J., and Zwarts, M. J. (2003). Analysis of recovery processes after stroke by means of transcranial magnetic stimulation. *J. Clin. Neurophysiol.* 20, 188–195. doi: 10.1097/00004691-200305000-00004

Hömberg, V., Stephan, K. M., and Netz, J. (1991). Transcranial stimulation of motor cortex in upper motor neurone syndrome: its relation to the motor deficit. *Electroencephalogr. Clin. Neurophysiol.* 81, 377–388. doi: 10.1016/0168-5597(91)90027-u

Hortobágyi, T., and Devita, P. (2006). Mechanisms responsible for the age-associated increase in coactivation of antagonist muscles. *Exerc. Sport Sci. Rev.* 34, 29–35. doi: 10.1097/00003677-200601000-00007

Hu, X. L., Tong, K. Y., Li, R., Xue, J. J., Ho, S. K., and Chen, P. (2012). The effects of electromechanical wrist robot assistive system with neuromuscular electrical stimulation for stroke rehabilitation. *J. Electromyogr. Kinesiol.* 22, 431–439. doi: 10.1016/j.jelekin.2011.12.010

Hu, X. L., Tong, K. Y., Song, R., Zheng, X. J., and Leung, W. W. (2009). A comparison between electromyography-driven robot and passive motion device on wrist rehabilitation for chronic stroke. *Neurorehabil. Neural Repair.* 23, 837–846. doi: 10.1177/1545968309338191

Hu, X. L., Tong, K. Y., Wei, X. J., Rong, W., Susanto, E. A., and Ho, S. K. (2013). The effects of post-stroke upper-limb training with an electromyography (EMG)-driven hand robot. *J. Electromyogr. Kinesiol.* 23, 1065–1074. doi: 10.1016/j.jelekin.2013.07.007

Hu, X., Suresh, A. K., Rymer, W. Z., and Suresh, N. L. (2015). Assessing altered motor unit recruitment patterns in paretic muscles of stroke survivors using surface electromyography. *J. Neural Eng.* 12:066001. doi: 10.1088/1741-2560/12/6/066001

Hu, X., Tong, K. Y., Song, R., Tsang, V. S., Leung, P. O., and Li, L. (2007). Variation of muscle coactivation patterns in chronic stroke during robot-assisted elbow training. *Arch. Phys. Med. Rehabil.* 88, 1022–1029. doi: 10.1016/j.apmr.2007.05.006

Jang, S. H. (2013). Motor recovery by improvement of limb-kinetic apraxia in a chronic stroke patient. *NeuroRehabilitation* 33, 195–200. doi: 10.3233/NRE-130945

Kamper, D. G., and Rymer, W. Z. (2001). Impairment of voluntary control of finger motion following stroke: role of inappropriate muscle coactivation. *Muscle Nerve* 24, 673–681. doi: 10.1002/mus.1054

Katz, R., and Pierrot-Deseilligny, E. (1982). Recurrent inhibition of alpha-motoneurons in patients with upper motor neuron lesions. *Brain* 105, 103–124. doi: 10.1093/brain/105.1.103

Kim, J. M., Tay, M. R. J., Rajeswaran, D. K., Tham, S. L., Lui, W. L., and Kong, K. H. (2021). Changes in muscle architecture on ultrasound in patients early after stroke. *NeuroRehabilitation* 49, 565–572. doi: 10.3233/NRE-210257

Kobayashi, M., and Pascual-Leone, A. (2003). Transcranial magnetic stimulation in neurology. *Lancet Neurol.* 2, 145–156.

Langhorne, P., Coupar, F., and Pollock, A. (2009). Motor recovery after stroke: a systematic review. *Lancet Neurol.* 8, 741–754. doi: 10.1016/s1474-4422(09)70150-4

Lee, J. A., Hwang, P. W., and Kim, E. J. (2015). Upper extremity muscle activation during drinking from a glass in subjects with chronic stroke. *J. Phys. Ther. Sci.* 27, 701–703. doi: 10.1589/jpts.27.701

Lemon, R. N. (2008). Descending pathways in motor control. *Annu. Rev. Neurosci.* 31, 195–218. doi: 10.1146/annurev.neuro.31.060407.125547

Li, G., Shourijeh, M. S., Ao, D., Patten, C., and Fregly, B. J. (2020). How well do commonly used co-contraction indices approximate lower limb joint stiffness trends during gait for individuals post-stroke? *Front. Bioeng. Biotechnol.* 8:588908. doi: 10.3389/fbioe.2020.588908

Madhavan, S., Krishnan, C., Jayaraman, A., Rymer, W. Z., and Stinear, J. W. (2011). Corticospinal tract integrity correlates with knee extensor weakness in chronic stroke survivors. *Clin. Neurophysiol.* 122, 1588–1594. doi: 10.1016/j.clinph.2011.01.011

Mang, C. S., Lagerquist, O., and Collins, D. F. (2010). Changes in corticospinal excitability evoked by common peroneal nerve stimulation depend on stimulation frequency. *Exp. Brain Res.* 203, 11–20. doi: 10.1007/s00221-010-2202-x

McMorland, A. J., Runnalls, K. D., and Byblow, W. D. (2015). A neuroanatomical framework for upper limb synergies after stroke. *Front. Hum. Neurosci.* 9:82. doi: 10.3389/fnhum.2015.00082

Misra, U. K., and Kalita, J. (1995). Motor evoked potential changes in ischaemic stroke depend on stroke location. *J. Neurol. Sci.* 134, 67–72. doi: 10.1016/0022-510x(95)00216-4

Morita, H., Shindo, M., Momoi, H., Yanagawa, S., Ikeda, S., and Yanagisawa, N. (2006). Lack of modulation of Ib inhibition during antagonist contraction in spasticity. *Neurology* 67, 52–56. doi: 10.1212/01.wnl.0000223399.59212.f4

Moritani, T. (1993). Neuromuscular adaptations during the acquisition of muscle strength, power and motor tasks. *J. Biomech.* 26 Suppl 1, 95–107. doi: 10.1016/0021-9290(93)90082-p

Nakayama, H., Jørgensen, H. S., Raaschou, H. O., and Olsen, T. S. (1994). Recovery of upper extremity function in stroke patients: the copenhagen stroke study. *Arch. Phys. Med. Rehabil.* 75, 394–398. doi: 10.1016/0003-9993(94)90161-9

Nam, C., Rong, W., Li, W., Xie, Y., Hu, X., and Zheng, Y. (2017). The effects of upper-limb training assisted with an electromyography-driven neuromuscular electrical stimulation robotic hand on chronic stroke. *Front. Neurol.* 8:679. doi: 10.3389/fneur.2017.00679

Ohn, S. H., Yoo, W. K., Kim, D. Y., Ahn, S., Jung, B., Choi, I., et al. (2013). Measurement of synergy and spasticity during functional movement of the post-stoke hemiplegic upper limb. *J. Electromyogr. Kinesiol.* 23, 501–507. doi: 10.1016/j.jelekin.2012.10.001

Okamoto, Y., Ishii, D., Yamamoto, S., Ishibashi, K., Wakatabi, M., Kohno, Y., et al. (2021). Relationship between motor function, DTI, and neurophysiological parameters in patients with stroke in the recovery rehabilitation unit. *J. Stroke Cerebrovasc. Dis.* 30:105889. doi: 10.1016/j.jstrokecerebrovasdis.2021.105889

Pan, B., Huang, Z., Jin, T., Wu, J., Zhang, Z., and Shen, Y. (2021). Motor function assessment of upper limb in stroke patients. *J. Healthc Eng.* 2021:6621950.

Pennisi, G., Alagona, G., Rapisarda, G., Nicoletti, F., Costanzo, E., Ferri, R., et al. (2002). Transcranial magnetic stimulation after pure motor stroke. *Clin. Neurophysiol.* 113, 1536–1543. doi: 10.1016/s1388-2457(02)00255-9

Potter-Baker, K. A., Varnerin, N. M., Cunningham, D. A., Roelle, S. M., Sankarasubramanian, V., Bonnett, C. E., et al. (2016). Influence of corticospinal tracts from higher order motor cortices on recruitment curve properties in stroke. *Front. Neurosci.* 10:79. doi: 10.3389/fnins.2016.00079

Qian, Q., Hu, X., Lai, Q., Ng, S. C., Zheng, Y., and Poon, W. (2017). Early stroke rehabilitation of the upper limb assisted with an electromyography-driven neuromuscular electrical stimulation-robotic arm. *Front. Neurol.* 8:447. doi: 10.3389/fneur.2017.00447

Ridding, M. C., Mckay, D. R., Thompson, P. D., and Miles, T. S. (2001). Changes in corticomotor representations induced by prolonged peripheral nerve stimulation in humans. *Clin. Neurophysiol.* 112, 1461–1469. doi: 10.1016/s1388-2457(01)00592-2

Roger, V. L., Go, A. S., Lloyd-Jones, D. M., Benjamin, E. J., Berry, J. D., Borden, W. B., et al. (2012). Heart disease and stroke statistics–2012 update: a report from the American Heart Association. *Circulation* 125, e2–e220.

Rong, W., Li, W., Pang, M., Hu, J., Wei, X., Yang, B., et al. (2017). A Neuromuscular Electrical Stimulation (NMES) and robot hybrid system for multi-joint coordinated upper limb rehabilitation after stroke. *J. Neuroeng. Rehabil.* 14:34. doi: 10.1186/s12984-017-0245-y

Rong, W., Tong, K. Y., Hu, X. L., and Ho, S. K. (2015). Effects of electromyography-driven robot-aided hand training with neuromuscular electrical stimulation on hand control performance after chronic stroke. *Disabil. Rehabil. Assist. Technol.* 10, 149–159. doi: 10.3109/17483107.2013.873491

Rossini, P. M., and Rossi, S. (2007). Transcranial magnetic stimulation: diagnostic, therapeutic, and research potential. *Neurology* 68, 484–488. doi: 10.1212/01.wnl.0000250268.13789.b2

Rossini, P. M., Barker, A. T., Berardelli, A., Caramia, M. D., Caruso, G., Cracco, R. Q., et al. (1994). Non-invasive electrical and magnetic stimulation of the brain, spinal cord and roots: basic principles and procedures for routine clinical application. Report of an IFCN committee. *Electroencephalogr. Clin. Neurophysiol.* 91, 79–92. doi: 10.1016/0013-4694(94)90029-9

Rossini, P. M., Burke, D., Chen, R., Cohen, L. G., Daskalakis, Z., Di Iorio, R., et al. (2015). Non-invasive electrical and magnetic stimulation of the brain, spinal cord, roots and peripheral nerves: basic principles and procedures for routine clinical and research application. An updated report from an I.F.C.N. Committee. *Clin. Neurophysiol.* 126, 1071–1107. doi: 10.1016/j.clinph.2015.02.001

Safavynia, S. A., Torres-Oviedo, G., and Ting, L. H. (2011). Muscle synergies: implications for clinical evaluation and rehabilitation of movement. *Top. Spinal Cord. Inj. Rehabil.* 17, 16–24. doi: 10.1310/sci1701-16

Sarcher, A., Raison, M., Ballaz, L., Lemay, M., Leboeuf, F., Trudel, K., et al. (2015). Impact of muscle activation on ranges of motion during active elbow movement in children with spastic hemiplegic cerebral palsy. *Clin. Biomech. (Bristol, Avon)* 30, 86–94. doi: 10.1016/j.clinbiomech.2014.10.009

Shimomura, Y., Iwanaga, K., Harada, H., and Katsuura, T. (1999). Evaluation and design of a small portable EMG amplifier with potential RMS output. *Appl. Hum. Sci.* 18, 61–67. doi: 10.2114/jpa.18.61

Silva, C. C., Silva, A., Sousa, A., Pinheiro, A. R., Bourlinova, C., Silva, A., et al. (2014). Co-activation of upper limb muscles during reaching in post-stroke subjects: an analysis of the contralesional and ipsilesional limbs. *J. Electromyogr. Kinesiol.* 24, 731–738. doi: 10.1016/j.jelekin.2014.04.011

Song, R., and Tong, K. Y. (2013). EMG and kinematic analysis of sensorimotor control for patients after stroke using cyclic voluntary movement with visual feedback. *J. Neuroeng. Rehabil.* 10:18. doi: 10.1186/1743-0003-10-18

Souissi, H., Zory, R., Bredin, J., Roche, N., and Gerus, P. (2018). Co-contraction around the knee and the ankle joints during post-stroke gait. *Eur. J. Phys. Rehabil. Med.* 54, 380–387. doi: 10.23736/S1973-9087.17.04722-0

Tanzarella, S., Muceli, S., Del Vecchio, A., Casolo, A., and Farina, D. (2020). Non-invasive analysis of motor neurons controlling the intrinsic and extrinsic muscles of the hand. *J. Neural Eng.* 17:046033. doi: 10.1088/1741-2552/aba6db

Traversa, R., Cicinelli, P., Bassi, A., Rossini, P. M., and Bernardi, G. (1997). Mapping of motor cortical reorganization after stroke. A brain stimulation study with focal magnetic pulses. *Stroke* 28, 110–117. doi: 10.1161/01.str.28.1.110

Traversa, R., Cicinelli, P., Pasqualetti, P., Filippi, M., and Rossini, P. M. (1998). Follow-up of interhemispheric differences of motor evoked potentials from the 'affected' and 'unaffected' hemispheres in human stroke. *Brain Res.* 803, 1–8. doi: 10.1016/s0006-8993(98)00505-8

Turton, A., Wroe, S., Trepte, N., Fraser, C., and Lemon, R. N. (1996). Contralateral and ipsilateral EMG responses to transcranial magnetic stimulation during recovery of arm and hand function after stroke. *Electroencephalogr. Clin. Neurophysiol.* 101, 316–328. doi: 10.1016/0924-980x(96)95560-5

Van Wittenberghe, I. C., and Peterson, D. C. (2021). *Corticospinal Tract Lesion. StatPearls*. Treasure Island, FL: StatPearls Publishing LLC.

Wagner, J. M., Dromerick, A. W., Sahrmann, S. A., and Lang, C. E. (2007). Upper extremity muscle activation during recovery of reaching in subjects with post-stroke hemiparesis. *Clin. Neurophysiol.* 118, 164–176. doi: 10.1016/j.clinph.2006.09.022

Wolfe, C. D. (2000). The impact of stroke. *Br. Med. Bull.* 56, 275–286.

Ziemann, U., Reis, J., Schwenkreis, P., Rosanova, M., Strafella, A., Badawy, R., et al. (2015). TMS and drugs revisited 2014. *Clin. Neurophysiol.* 126, 1847–1868. doi: 10.1016/j.clinph.2014.08.028

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 | Frontiers in Neuroscience

ORIGINAL RESEARCH
published: 07 June 2022
doi: 10.3389/fnins.2022.900146

Flexible Non-contact Electrodes for Wearable Biosensors System on Electrocardiogram Monitoring in Motion

Xin Wang[1,2,3†], Shuting Liu[4†], Mingxing Zhu[1,5], Yuchao He[1,3], Zhilong Wei[1,3], Yingying Wang[1,2,3], Yangjie Xu[6], Hongguang Pan[7], Weimin Huang[8], Shixiong Chen[1,3*] and Guanglin Li[1,3]*

[1] CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, [2] Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen, China, [3] Guangdong-Hong Kong-Macao Joint Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, [4] Department of Informatics, Technical University of Munich, Munich, Germany, [5] School of Electronics and Information Engineering, Harbin Institute of Technology, Shenzhen, China, [6] Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, Luxembourg, Luxembourg, [7] Department of Otolaryngology, Shenzhen Children's Hospital, Shenzhen, China, [8] Department of Neonatology, Shenzhen Children's Hospital, Shenzhen, China

Edited by:
Weidong Gao,
Beijing University of Posts and Telecommunications (BUPT), China

Reviewed by:
Lisheng Xu,
Northeastern University, China
Wang Mingjiang,
Harbin Institute of Technology, China

***Correspondence:**
Mingxing Zhu
458260329@qq.com
Shixiong Chen
sx.chen@siat.ac.cn

† These authors have contributed equally to this work

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

Received: 20 March 2022
Accepted: 10 May 2022
Published: 07 June 2022

Citation:
Wang X, Liu S, Zhu M, He Y, Wei Z, Wang Y, Xu Y, Pan H, Huang W, Chen S and Li G (2022) Flexible Non-contact Electrodes for Wearable Biosensors System on Electrocardiogram Monitoring in Motion. Front. Neurosci. 16:900146. doi: 10.3389/fnins.2022.900146

Electrocardiogram (ECG) is a critical physiological indicator that contains abundant information about human heart activities. However, it is a kind of weak low-frequency signal, which is easy to be interfered by various noises. Therefore, wearable biosensors (WBS) technique is introduced to overcome this challenge. A flexible non-contact electrode is proposed for wearable biosensors (WBS) system, which is made up of flexible printed circuits materials, and can monitor the ECG signals during exercise for a long time. It uses the principle of capacitive coupling to obtain high-quality signals, and reduces the impact of external noise through active shielding; The results showed that the proposed non-contact electrode was equivalent to a medical wet electrode. The correlation coefficient was as high as 99.70 ± 0.30% when the subject was resting, while it was as high as 97.53 ± 1.80% during exercise. High-quality ECG could still be collected at subjects walking at 7 km/h. This study suggested that the proposed flexible non-contact electrode would be a potential tool for wearable biosensors for medical application on long-term monitoring of patients' health and provide athletes with physiological signal measurements.

Keywords: wearable biosensors system, flexible non-contact electrode, electrocardiogram, monitoring, motion

INTRODUCTION

Physiological signals are widely used in medical applications and health monitoring, and they serve as the objective indicators closely related to human health (Le et al., 2010). For example, the ECG signals that contain abundant information of the heart activities provide useful information for disease prevention (Zhang et al., 1997). And it is a critical physiological indicator for screening cardiovascular diseases and evaluating heart or cardiovascular functions. Therefore, the standard wet silver chloride (Ag/AgCl) electrodes as the most commonly used biopotential electrodes for measuring electrocardiogram (ECG) signals is found in hospital-based medical diagnostic and home health monitoring systems (Fernandez and Pallas-Areny, 2000). However, the ECG signal is a low-frequency weak signal and is susceptible to various noise interferences during acquisition (Song and Yu, 2010). For example, Muscle cells tremble as they contract, producing high-frequency EMG signals, which have a high frequency and a low amplitude (Deng et al., 2000). This EMG signal will cause interference to the original signal when collecting ECG. The affected ECG signal appears as a small glitch in the time domain. In addition, breathing can also affect the acquisition of signals. While breathing, people are accompanied by small fluctuations in the body, producing lower frequency interference signals (Wu et al., 2014). The baseline of a normal ECG signal is a straight line, while the baseline of the interfered signal is no longer a straight line, often referred to as baseline drift (Chouhan and Mehta, 2007; Yu et al., 2014). Moreover, the impact of motion on signal acquisition is more obvious (Garcia-Casado et al., 2006; Lee et al., 2014; Torfs et al., 2014; Pei et al., 2016). This kind of interference is uncontrollable and will cause a large change in the signal. Therefore, general ECG acquisition requires the subject to remain as static as possible (Liu et al., 2018). In the state of motion, the electrode and the human skin will be distorted and displaced. Friction can cause changes in the impedance between the skin and the electrodes, causing some interference during the entire process.

Due to the special occupation of athletes, higher requirements are placed on the electrodes. Currently used in the clinic is a standard wet electrode, which reduces the impedance between the skin and the electrode through a conductive gel to achieve high-quality ECG (Li et al., 2018). However, there are still lots of problems to be solved. For high-quality ECG monitoring, researchers did their efforts on the Wearable Biosensors, most of which are dry electrodes that do not require conductive gels (Dozio et al., 2007; Assambo and Burke, 2009; Zhang et al., 2015; Joutsen et al., 2017). Kim et al. (2016) proposed a 1D-2D hybrid carbon nanocomposites-based ECG electrode and recorded ECG with three movements, namely wrist curl, squat, and writing. Noh et al. (2016) proposed a novel conductive carbon black and PDMS ECG electrode which could achieve ECG recording with water exposed conditions. Asadi et al. (2021) realized continuous medical ECG monitoring by constructing a graphene elastomer electrode. Even though plenty of WBSs were proposed, none of the existing methods met the requirement on ECG monitoring in motion. Besides, the dry electrodes without conductive gel would have the impedance between electrodes and skin rather high, which makes the recorded signal be easily affected by the body motion of the subjects (Pei et al., 2016). Non-contact electrodes, which can make measurements through clothing without any contact between the electrode and the skin, is essential for building user-friendly WBS networks for physiological recording during motion (Lee et al., 2013). Up to date, most of the non-contact electrodes are rigid and sensitive to motion (Chi et al., 2009, 2010, 2011, 2013), not benefiting for long-term monitoring. Therefore, monitoring ECG signals during motion is definitely a challenge.

In this study, a flexible non-contact electrode for WBS was proposed based on the principle of capacitive coupling for physiological signal acquisition. The electrode was built with flexible printed circuits (FPC) materials and could be bent to ensure better capacitive coupling with the skin (Yi et al., 2018; Liu et al., 2019). Although the motion had a great influence, the proposed non-contact electrode could acquire high-quality ECG compared with the wet electrode. In this study, we first studied the effects of the layers of the insulation materials on flexible non-contact electrode-based ECG quality compared with the gold standard method. Then, we further studied on the effect of the size of the flexible non-contact electrode on the ECG quality. After that, we investigated the performance of flexible non-contact electrodes on ECG recording at different walking speeds.

MATERIALS AND METHODS

Subjects

Three male subjects aged from 20 to 25 years old were recruited in this study. The subjects had normal cardiac function, normal muscle function, and no cognitive impairments. The experimental procedures were clearly explained to the subjects, and the data collection was carried out in an ordinary laboratory environment without any electromagnetic shielding. All the experimental protocols were approved by the Institutional Review Board (IRB) of the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. (SIAT-IRB-190615-H0352).

The Experimental Scheme

The overall design block diagram of the system was shown in **Figure 1**. Two non-contact electrode sheets and one driven right leg (DRL) electrode were used to collect physiological electrical signals (Guermandi et al., 2008; Kim and Park, 2008; Guerrero and Spinelli, 2017). The physiological signal monitoring of non-contact electrodes mainly utilizes the principle of capacitive coupling. The non-contact electrode sheet and the skin corresponds to both the conductive surface of the capacitor, while the clothing corresponds to the insulating dielectric-filled in the middle of the capacitor. The signal coupled into the circuit was pre-processed by an anti-aliasing filter, and then entered the ADS1299 (Texas Instruments, Dallas, Texas, United States), a front-end board, for signal acquisition and conversion from analog signals into high-resolution digital signals. The data acquisition process was controlled by the wireless CC3200 MCU (Texas Instruments, Dallas, Texas, United States) and the data from ADS1299 was streamed through high-speed Wi-Fi to the PC, where a Matlab (Mathworks Inc., United States) GUI was ready to display the real-time waveform and stored the raw data for offline analyses. The physiological electrical signal acquisition system was shown in **Figure 2**, including a circuit board, electrode board, and electrode sheets. The electrode sheet was of round shape and composed of a double-layer flexible printed circuit board (FPCB), which could be

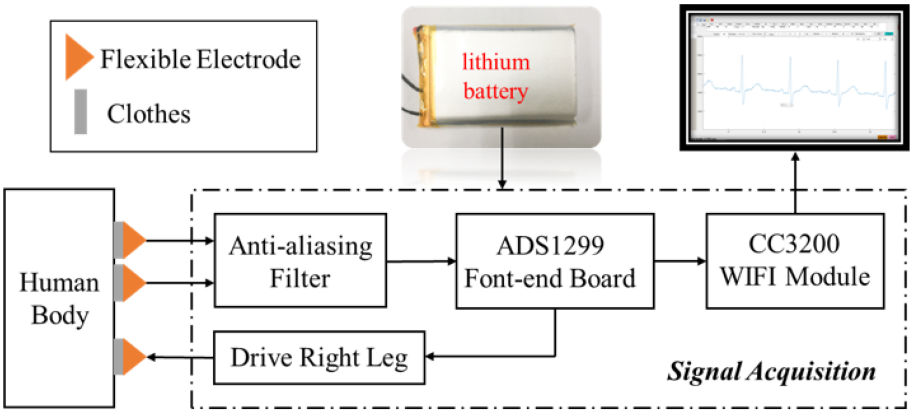


FIGURE 1 | The overall design block diagram of the non-contact ECG signal acquisition system.

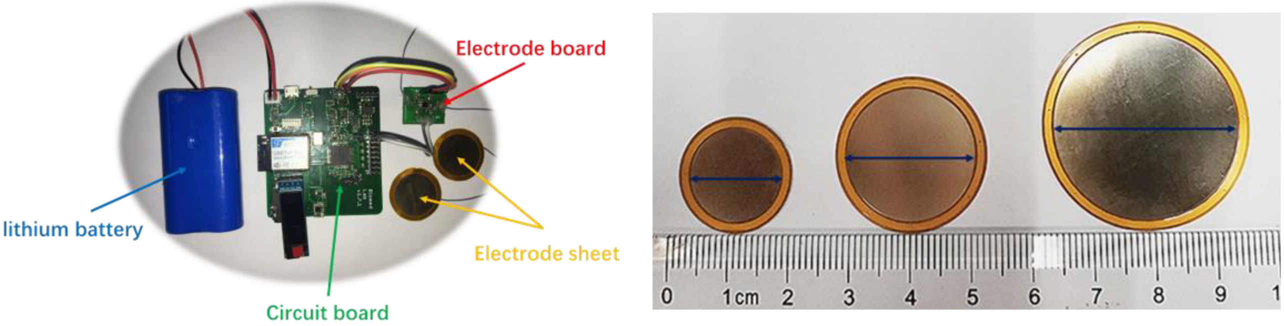


FIGURE 2 | The photos of the ECG signal acquisition system and flexible non-contact electrodes with different sizes [small (the diameter is 1.5 cm), medium (the diameter is 2.1 cm), and large (the diameter is 3.0 cm)] electrode.

bent according to the skin curve. The bottom layer was completely filled with copper to prevent the sensing plate from external interferences; the top layer was composed of one circular copper-filled non-contact capacitive sensing plate and one concentric outer shielding ring connected with the outer ring at the bottom to improve the shielding outcomes (Muneer, 2014; Chen et al., 2017; Jiang et al., 2017). The flexible non-contact electrode was connected with the electrode board using a soft shielded cable, whose inner wire was connected with the top sensing plate and the outer shield was connected with the shielding layer. The ADS1299 and CC3200 WIFI modules were soldered on the circuit board. In addition, the lithium battery was used for the power supply to reduce power frequency interference.

RESULTS

The Effects of Insulation Layers and Electrode Size

To verify the performance of the proposed electrodes, the standard wet electrodes were used as a reference, which was shown by the blue line in **Figure 3**. The red signals were the ECG signals collected by our electrodes when a layer of canvas was used as the insulating material, while the pink signal was acquired when three layers of canvas were used as the insulating material. All signals were collected synchronously while the subject remained relaxed stationary position. It was clear that the ECG signal quality was high, the baseline was very thin and stable and P-wave, T-wave, and QRS waves could be clearly observed. However, as the number of interlayer materials increased, the signal quality tended to decline. In the figure, the baseline was widened and the signal contains burrs, but it did not evidently influence the quality. The second, third, and fourth rows in the figure used small (the diameter is 1.5 cm), medium (the diameter is 2.1 cm), and large (the diameter is 3.0 cm) electrode sheets, respectively. As the size of the electrode sheets increased, the amplitude of the ECG signal increased a little bit.

Among them, the medium electrode signal quality was the best, and the electrode sheet was too large or too small would have an impact on the signal quality. If the electrode piece was too small, it would cause a less effective signal to be transmitted, affecting the signal quality, while if the electrode piece was too large, and in the case of non-contact, the noise generated by the friction would be mixed into the signal, reducing the signal-to-noise ratio. In all, our non-contact electrodes could acquire high-quality ECG signals which were

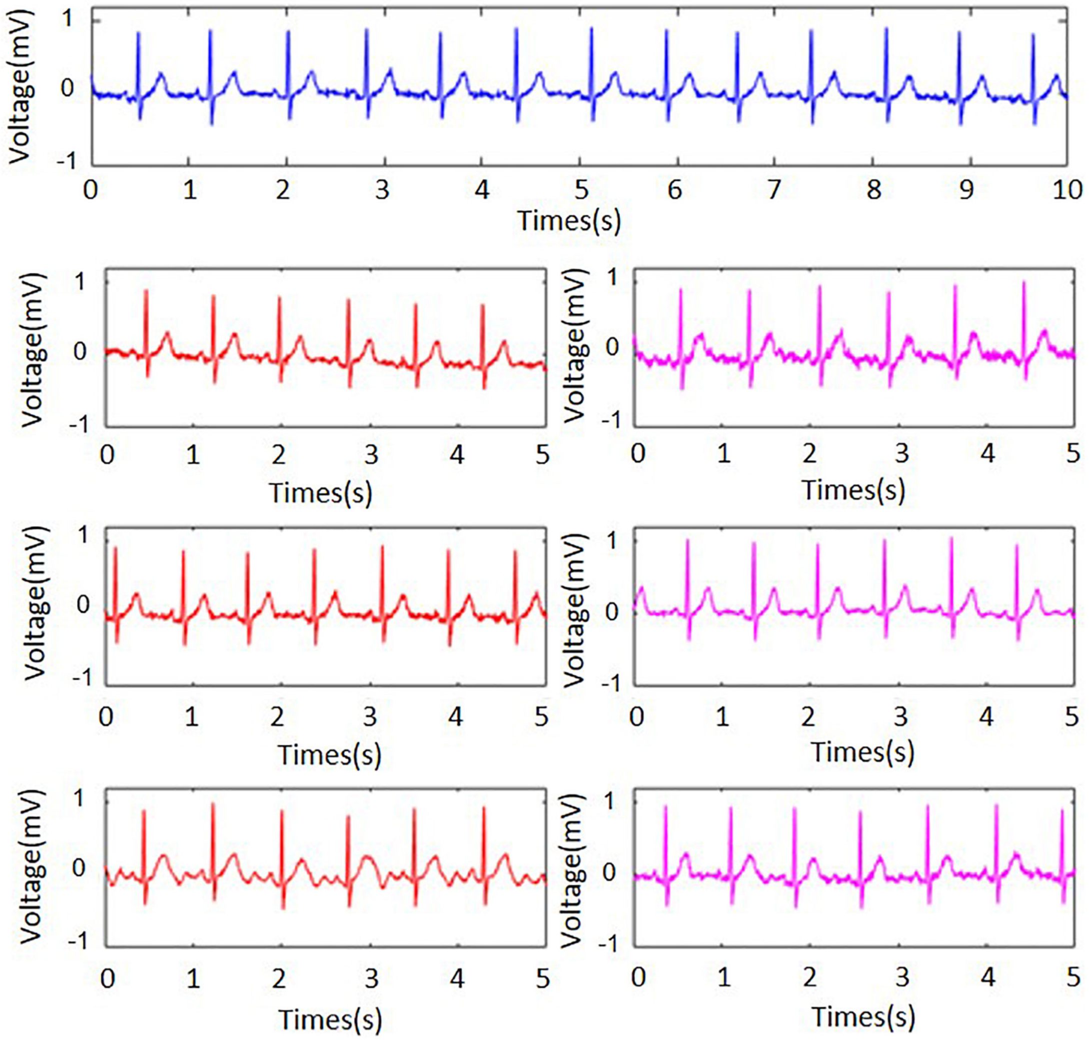


FIGURE 3 | Static electrocardiographic signals collected under different electrodes (blue for wet electrode, red and pink for flexible non-contact electrode) and under different thicknesses of canvas (red for a layer of canvas and pink for three layers). The second, third, and fourth rows used small, medium, and large electrodes, respectively.

highly similar to the standard wet electrode signal, when the subject was quiet.

Electrocardiogram Collected During Motion

The flexible non-contact electrodes could not only capture signals when the subject was quiet but also acquired signals when they were in motion. When the subject was exercising at a speed of 2 km/h on the treadmill, we selected the medium electrode piece as the non-contact electrode for signal monitoring. At the same time, we used the standard wet electrode and the dry electrode for simultaneous monitoring, and the positions of the electrode pieces were as close as possible. The original ECG time-domain waveforms obtained by the three different electrodes including the wet, flexible non-contact, and dry electrodes, were shown in **Figure 4**. It could be observed that the signals collected by the three electrodes were highly similar. Therefore, although the subject was in motion, the proposed non-contact electrodes, especially the flexible one, could achieve a comparable quality of signals with the standard wet electrodes in ECG measurements.

Electrocardiogram Collected With Different Walking Speeds

Next, we used the flexible non-contact electrode to further investigate the ECG monitoring during exercise. Subjects were kept at a constant speed from 2 to 7 km/h on the treadmill. At the same time, we used non-contact electrodes and standard wet electrodes to monitor ECG signals, as shown in **Figure 5**. It could be observed from the figure that as the speed of the subject increased, the number of R waves detected during the same time increased, which was in line with the mechanism of human motion. Besides, as the speed

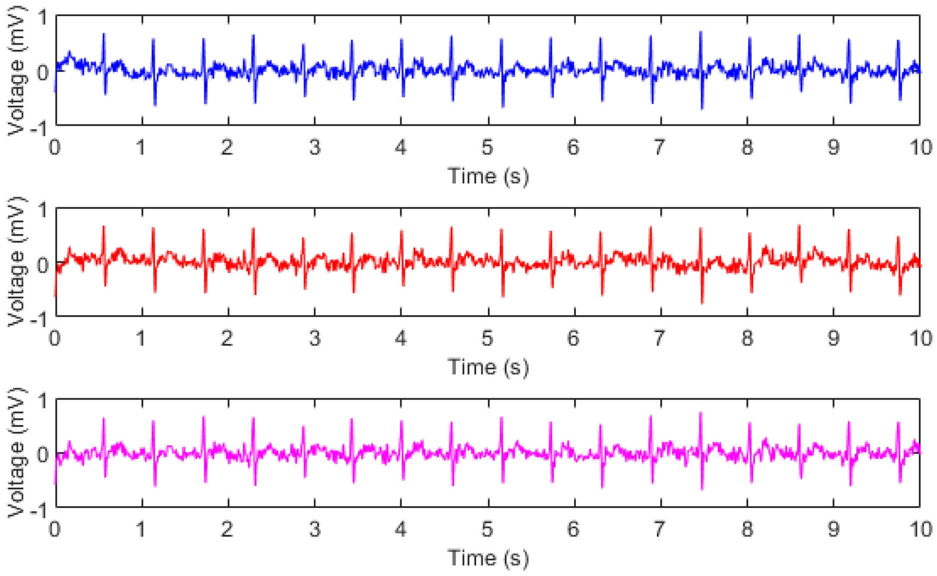


FIGURE 4 | The ECG recorded by wet electrode (blue), proposed electrode (red), dry electrode (pink) at the speed of 2 km/h.

increased, the signal quality of both the non-contact electrode and the wet electrode decreased, and the T wave and the P wave were gradually unclear, but the R wave could still be clearly detected.

Mathematical Statistics Analysis

When the subject remained stationary, the small, medium, and large electrodes all had a high correlation coefficient with the wet electrode, as shown in **Figure 6**, indicating that the perfectly designed flexible electrode can perform ECG monitoring in a non-contact state. When studying the effects of insulation layers and electrode size, the ECG signals from the wet and non-contact electrodes were compared and analyzed by calculating the average correlation coefficient. An average correlation coefficient of 99.28 ± 0.42, 99.70 ± 0.30, and 97.18 ± 0.52% was calculated for the small, medium and large ECG electrodes, respectively when compared to the wet Ag/AgCl electrode. Therefore, the size of the electrode would bring about a difference in the correlation coefficient, which meant that the size of the electrode would affect the signal. For example, the medium-sized electrode had the best signal. When the electrode pad area became smaller, the effective signal would decrease; when the electrode pad area became larger, the noise entering the circuit would also increase. When considering the condition in motion, taking the standard wet electrode as the gold standard, the correlation coefficient between the non-contact electrode and the wet electrode was as high as 97.53 ± 1.80%, and the correlation coefficient between the dry electrode and the wet electrode was as high as 98.43 ± 0.77%. Moreover, we calculated the SNR of the ECG acquired by the proposed electrode under different walking speeds, the result showed that with the increment of the speeds, the quality of the ECG could maintain reliable.

DISCUSSION

Effects of Electrode Size

The non-contact electrode uses the principle of capacitive coupling (Wang et al., 2015; Li and Sun, 2017). According to the formula of the capacitance, the capacitance increases as the area of the electrode sheet increases, and decreases as the distance between the electrode sheet and the skin increases. However, according to actual experiments, the quality and strength of effective signals do not increase with the increase in electrode pad size. In this experiment, we used non-contact electrodes with diameters of 1.5, 2.1, and 3 cm, with an area ratio of 1:1.4:2, respectively. As the area of the electrode sheets increased, the amplitude of the electrocardiographic signal appeared to become larger, but not significant. At the same time, we used the method of calculating the correlation coefficient to evaluate the ECG signal quality compared with the standard wet electrode. An average correlation coefficient of 99.28 ± 0.42, 99.70 ± 0.30, and 97.18 ± 0.52% was calculated for the small, medium and large ECG electrodes, respectively when compared to the wet Ag/AgCl electrode. The results demonstrated that the proposed electrode performed as better as the gold standard method does. Besides, we found that the signal quality of the medium electrode was the best, so in the non-contact acquisition, the quality of the ECG signal and the area of the electrode pad were not simply proportional. It was possible that as the area of the electrode sheet became larger, the effective signal was increased, but at the same time, noise interference was also introduced.

Effects of Clothing Thickness

Since the non-contact electrode is based on the capacitive coupling between the electrode plate and the skin surface, the

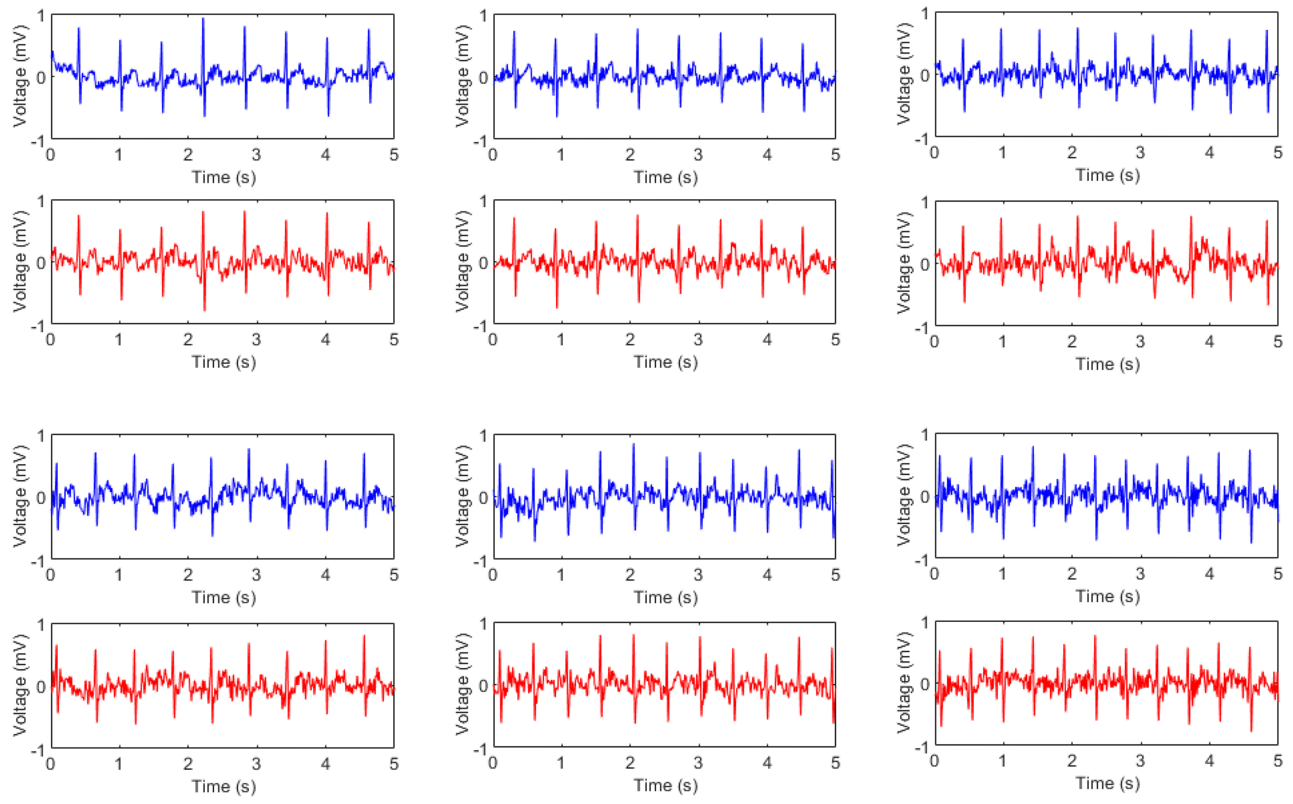


FIGURE 5 | The ECG signals recorded by wet (blue) and flexible non-contact electrode (red) under different walking speeds (from left to right and from up to down, they were 2, 3, 4, 5, 6, and 7 km/h, respectively).

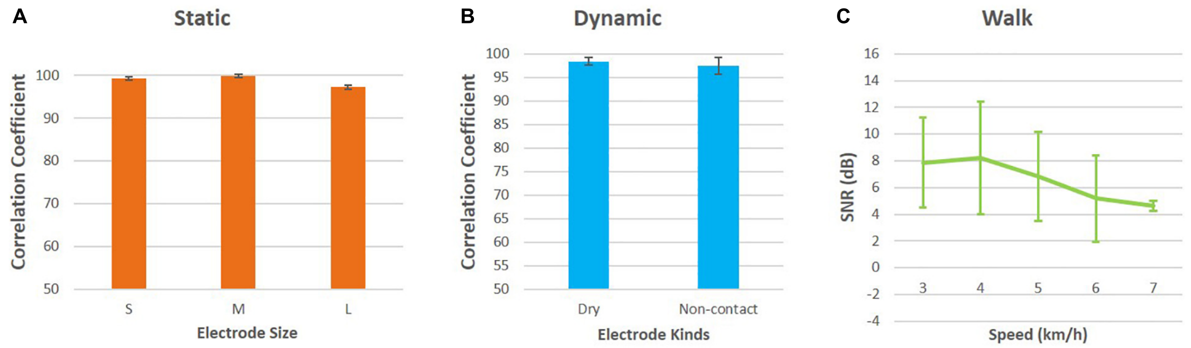


FIGURE 6 | Mathematical statistics analysis: **(A)** Correlation coefficients between electrodes of different sizes and wet electrodes under static conditions; **(B)** correlation coefficients between flexible non-contact electrodes, flexible dry electrodes and wet electrodes under motion; **(C)** SNR at different walking speeds.

effective capacitance is proportional to the area of the electrode plate and the dielectric constant of the insulating material while inversely proportional to the distance between the electrode and the skin. As the thickness of the insulating layer increases, the signal quality decreases. In theory, the further the electrode is from the source, the smaller the coupling signal (Liu et al., 2019). In this experiment, we used two thicknesses of clothing for monitoring and both obtained an effective ECG signal, namely one layer and three layers. The result showed that thin clothes had less noise interference than thick clothes. It was consistent with the principle of capacitive coupling. As the layers of the insulation materials increase, the distance between the electrode and the skin increase, which finally causes the decrease of the effective capacitance and therefore leads to the deterioration of the ECG quality. This result was consistent with the existing research that had been reported (Chi et al., 2009; Mathias et al., 2015; Wang et al., 2019). And in our study, for three layers of the insulating material, the ECG signals still had recognizable ECG characteristics, which could be used for further analysis.

Effects of Walking Speed

We not only performed ECG signal acquisition when the subject was at rest but also monitored the subject at constant speeds ranging from 2 to 7 km/h with the increment of 1 km/h. The results showed that as the speed increased, the ECG cycle became shorter, which was consistent with the mechanism of human motion that the heart would beat faster when the human being did exercise. As the speed increased, the wet-electrode and our electrode signal quality tended to decrease, mainly affected by myoelectric signals and noise from electrodes friction and breathing. When the subject was walking, muscle cells tremble as they contract, producing high-frequency EMG signals, which had a high frequency and a low amplitude. This EMG signal would cause interference to the original signal when collecting ECG. The affected ECG signal appeared as a small glitch in the time domain. With the increase of speed, the EMG signal became stronger and the interference to the ECG became larger. During the exercise, the electrodes and skin would rub against the clothes and cause relative displacement. At this time, a large amount of motion artifact noise was generated. It was no doubt that motion artifacts had a great impact on the signal. As the speed increased, the motion artifacts also increased, therefore, leading to the quality of the ECG signal decreasing. In addition, breathing could also affect the acquisition of signals. While breathing, people were accompanied by small fluctuations in the body, producing lower frequency interference signals. The baseline of a normal ECG signal was a straight line, while the baseline of the interfered signal was no longer a straight line, often referred to as baseline drift. The deterioration of signal quality was mainly manifested by the thickening of the baseline and the unclear T and P waves. However, the QRS waveform could be clearly seen in this study. Overall, our electrodes were resistant to noise interference during motion monitoring, with good signal quality and a correlation coefficient of 97.53 ± 1.80% with standard wet electrodes. The SNR of ECG in 3, 4, 5, 6, and 7 km/h were 7.87 ± 3.36, 8.23 ± 4.22, 6.83 ± 3.33, 5.20 ± 3.24, and 4.64 ± 0.40 dB, respectively. This result suggested that the proposed electrode could achieve high quality ECG recording even in motion. In the future, we would like to recruit more subjects with different sexualities and weights to further evaluate the proposed electrode.

Noises in Electrocardiogram

In common practices, the recorded ECG signals are easily contaminated by noises such as powerline interferences, electrode displacement, electrode lead jitter, EMG noises and so on. It could be observed from **Figures 3–5** that the recorded ECG signals for our proposed non-contact electrode did now show significant contamination by the abovementioned noises. Such anti-noise capacity was achieved by the following technologies of our hardware and experimental protocol: (1) the active shielding technology (Jiang et al., 2018) in which all the frontend pathway of the non-contact electrode was shielded by a buffered voltage of the same level as the inner-wire signal, so that the capacitive coupling to the mains and to the ground could almost be eliminated, leading to extraordinary attenuation of the powerline interferences. With the assistance of the same active shielding technology, the interferences introduced by electrode lead jitter during body movements were also eliminated, which simplifies the subsequent signal processing and analysis tasks. (2) the proposed non-contact electrode was quite flexible so that it could bend freely according to the local curvation of the skin. In this way, it was very difficult for the electrode to move after it was fixed, even when the subject was running. The noises introduced by electrode displacement were prevented from degrading the ECG signal quality as a result. (3) in the experimental protocol, the electrodes were placed in the forearm positions where there were few muscles underneath. The subjects were also asked to avoid making large forearm muscle contractions when running. In this way, there were no significant EMG noise interferences in the ECG recording of this study. The data acquisition system of this study with flexible non-contact electrode could be a great tool to measure ECG signals with robust performance in noise immunity.

CONCLUSION

In this study, we proposed a wearable device for dynamic ECG monitoring, whose electrodes were flexible and non-contact, which was very friendly to subjects. When developing WBS devices, consideration should be given to user comfort, stability of long-term monitoring, and reliability of acquired signals. To enhance user comfort, we designed a flexible capacitive electrode that could bend along the curvature of the body surface. This capacitive coupling means that the ECG signal can be monitored while the subject is wearing the clothes, which avoids the possibility of allergy when using a wet electrode. We designed the electrode shielding ring to reduce noise interference and monitor the motion of the motion center. The results showed that the proposed electrodes could effectively resist the interference of noise. The ECG signal collected at rest reached the correlation coefficient of 99.70 ± 0.30% compared with the standard wet electrode. While, in the motion state, the correlation coefficient of the signal was as high as 97.53 ± 1.80% compared with the wet electrodes and dry electrodes. We verified the performance of the electrode by walking on a treadmill at a speed of up to 7 km/h. The results confirmed the feasibility of the developed system in daily life ECG monitoring, and suggested that the proposed flexible non-contact electrode would be a potential candidate for WBS systems on long-term healthcare monitoring for patients and physiological signal measurements for athletes.

DATA AVAILABILITY STATEMENT

The data used to support the findings of this study are available on reasonable request from the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board of Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences (SIAT-IRB-190615-H0352). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SC: conceptualization and supervision. SL, XW, HP, and WH: data curation. SC, GL, and MZ: funding acquisition. SL, XW, and YW: methodology. XW, YH, and ZW: validation. XW and YX: visualization. SL and XW: writing—original draft. XW and SC: writing—review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the National Natural Science Foundation of China (#81927804 and #62101538), the Shenzhen Governmental Basic Research Grant (#JCYJ20180507182241622), the Science and Technology Planning Project of Shenzhen (#JSGG20210713091808027 and #JSGG20211029095801002), and SIAT Innovation Program for Excellent Young Researchers (E1G027).

REFERENCES

Asadi, S., He, Z., Heydari, F., Li, D., Yuce, M. R., Alan, T., et al. (2021). Graphene Elastomer Electrodes for Medical Sensing Applications: combining High Sensitivity, Low Noise and Excellent Skin Compatibility to Enable Continuous Medical Monitoring. *IEEE Sensors J.* 21, 13967–13975. doi: 10.1109/jsen.2020.3003300

Assambo, C., and Burke, M. J. (2009). Amplifier input impedance in dry electrode ECG recording. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2009, 1774–1777. doi: 10.1109/IEMBS.2009.5333398

Chen, C.-C., Chang, W.-Y., and Xie, T. Y. (2017). "Shielded capacitive electrode with high noise immunity," in *Proceedings of the 2017 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, (Piscataway: IEEE), 157–158.

Chi, Y. M., Deiss, S. R., and Cauwenberghs, G. (2009). "Non-contact low power EEG/ECG electrode for high density wearable biopotential sensor networks," in *Proceedings of the 2009 Sixth International Workshop on Wearable and Implantable Body Sensor Networks*, (Piscataway: IEEE), 246–250.

Chi, Y. M., Maier, C., and Cauwenberghs, G. (2011). "Integrated ultra-high impedance front-end for non-contact biopotential sensing," in *Proceedings of the 2011 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, (Piscataway: IEEE), 456–459.

Chi, Y. M., Ng, P., and Cauwenberghs, G. (2013). Wireless noncontact ECG and EEG biopotential sensors. *ACM Transac. Embedded Comput. Syst.* 12, 1–19.

Chi, Y. M., Ng, P., Kang, E., Kang, J., Fang, J., Cauwenberghs, G., et al. (2010). Wireless non-contact cardiac and neural monitoring. *In Wireless Health* 2010, 15–23. doi: 10.3390/s22031010

Chouhan, V. S., and Mehta, S. S. (2007). "Total removal of baseline drift from ECG signal," in *Proceedings of the 2007 International Conference on Computing: Theory and Applications*, (Piscataway: IEEE), 512–515.

Deng, Y., Wolf, W., Schnell, R., and Appel, U. (2000). New aspects to event-synchronous cancellation of ECG interference: an application of the method in diaphragmatic EMG signals. *IEEE Transac. Biomed. Eng.* 47, 1177–1184. doi: 10.1109/10.867924

Dozio, R., Baba, A., Assambo, C., and Burke, M. J. (2007). Time based measurement of the impedance of the skin-electrode interface for dry electrode ECG recording. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2007, 5001–5004. doi: 10.1109/IEMBS.2007.4353463

Fernandez, M., and Pallas-Areny, R. (2000). Ag-AgCl electrode noise in high-resolution ECG measurements. *Biomed. Instrument. Technol.* 34, 125–130.

Garcia-Casado, J., Martinez-de-Juan, J. L., and Ponce, J. L. (2006). Adaptive filtering of ECG interference on surface EEnGs based on signal averaging. *Physiol. Measurement* 27:509. doi: 10.1088/0967-3334/27/6/005

Guermandi, M., Scarselli, E. F., and Guerrieri, R. (2008). A driving right leg circuit (DgRL) for improved common mode rejection in bio-potential acquisition systems. *IEEE Transac. Biomed. Circ. Syst.* 2015, 507–517. doi: 10.1109/TBCAS.2015.2446753

Guerrero, F. N., and Spinelli, E. (2017). High gain driven right leg circuit for dry electrode systems. *Med. Eng. Phys.* 39, 117–122. doi: 10.1016/j.medengphy.2016.11.005

Jiang, Y., Ji, N., Wang, H., Liu, X., Geng, Y., Li, P., et al. (2017). "Comparison of different shielding methods in acquisition of physiological signals," in *Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (Piscataway: IEEE), 2325–2328. doi: 10.1109/EMBC.2017.8037321

Jiang, Y., Samuel, O. W., Liu, X., Wang, X., Idowu, P. O., Li, P., et al. (2018). Effective biopotential signal acquisition: comparison of different shielded drive technologies. *Appl. Sci.* 8:276.

Joutsen, A. S., Kaappa, E. S., Karinsalo, T. J., and Vanhala, J. (2017). "Dry electrode sizes in recording ECG and heart rate in wearable applications," in *EMBEC & NBC 2017. EMBEC NBC 2017 2017. IFMBE Proceedings, vol 65*, eds H. Eskola, O. Väisänen, J. Viik, and J. Hyttinen (Singapore: Springer).

Kim, K. K., and Park, K. S. (2008). "Effective coupling impedance for power line interference in capacitive-coupled ECG measurement system," in *Proceedings of the 2008 International Conference on Information Technology and Applications in Biomedicine*, (Piscataway: IEEE), 256–258.

Kim, T., Park, J., Sohn, J., Cho, D., and Jeon, S. (2016). Bioinspired, Highly Stretchable, and Conductive Dry Adhesives Based on 1D-2D Hybrid Carbon Nanocomposites for All-in-One ECG Electrodes. *ACS Nano* 10, 4770–4778. doi: 10.1021/acsnano.6b01355

Le, T. Q., Bukkapatnam, S. T., Sangasoongsong, A., and Komanduri, R. (2010). "Towards virtual instruments for cardiovascular healthcare: Real-time modeling of cardiovascular dynamics using ECG signals," in *In Proceedings of the 2010 IEEE International Conference on Automation Science and Engineering*, (Piscataway: IEEE), 903–910.

Lee, J. S., Heo, J., Lee, W. K., Lim, Y. G., Kim, Y. H., Park, K. S., et al. (2014). Flexible capacitive electrodes for minimizing motion artifacts in ambulatory electrocardiograms. *Sensors* 14, 14732–14743. doi: 10.3390/s140814732

Lee, S. M., Kim, J. H., Byeon, H. J., Choi, Y. Y., Park, K. S., Lee, S.-H., et al. (2013). A capacitive, biocompatible and adhesive electrode for long-term and cap-free monitoring of EEG signals. *J. Neural Eng.* 10:036006. doi: 10.1088/1741-2560/10/3/036006

Li, G., Wang, S., and Duan, Y. Y. (2018). Towards conductive-gel-free electrodes: understanding the wet electrode, semi-dry electrode and dry electrode-skin interface impedance using electrochemical impedance spectroscopy fitting. *Sensors Actuat. Chem.* 277, 250–260.

Li, X., and Sun, Y. (2017). ". NCMB-button: A wearable non-contact system for long-term multiple biopotential monitoring," in *Proceedings of the IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, (Piscataway: IEEE), 348–355.

Liu, S., Liu, X., Jiang, Y., Wang, X., Huang, P., Wang, H., et al. (2018). "Flexible non-contact electrodes for bioelectrical signal monitoring," in *Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (Piscataway: IEEE), 4305–4308. doi: 10.1109/EMBC.2018.8513306

Liu, S., Zhu, M., Liu, X., Samuel, O. W., Wang, X., Huang, Z., et al. (2019). Flexible noncontact electrodes for comfortable monitoring of physiological signals. *Int. J. Adaptiv. Contr. Signal Proc.* 33, 1307–1318.

Mathias, D. N., Kim, S.-I., Park, J.-S., Joung, Y.-H., and Choi, W. S. (2015). Electrode Characteristics of Non-contact Electrocardiographic Measurement. *Transac. Electr. Electron. Materials* 16, 42–45. doi: 10.4313/teem.2015.16.1.42

Muneer, K. A. (2014). "Non contact ECG recording instrument for continuous cardiovascular monitoring," in *Proceedings of the IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, (Piscataway: IEEE), 269–272.

Noh, Y., Bales, J. R., Reyes, B. A., Molignano, J., Clement, A. L., Pins, G. D., et al. (2016). Novel Conductive Carbon Black and Polydimethlysiloxane ECG Electrode: a Comparison with Commercial Electrodes in Fresh, Chlorinated, and Salt Water. *Ann. Biomed. Eng.* 44, 2464–2479. doi: 10.1007/s10439-015-1528-8

Pei, W., Zhang, H., Wang, Y., Guo, X., Xing, X., Huang, Y., et al. (2016). Skin-potential variation insensitive dry electrodes for ECG recording. *IEEE Transac. Biomed. Eng.* 64, 463–470. doi: 10.1109/TBME.2016.2562702

Song, L., and Yu, F. (2010). "The time-frequency analysis of abnormal ECG signals," in *Life System Modeling and Intelligent Computing*, eds K. Li, L. Jia, X. Sun, M. Fei, and G. W. Irwin (Berlin, Heidelberg: Springer), 60–66.

Torfs, T., Chen, Y.-H., Kim, H., and Yazicioglu, R. F. (2014). Noncontact ECG recording system with real time capacitance measurement for motion artifact reduction. *IEEE Transac. Biomed. Circ. Syst.* 8, 617–625. doi: 10.1109/TBCAS.2014.2359053

Wang, J., Fujiwara, T., Kato, T., and Anzai, D. (2015). Wearable ECG based on impulse-radio-type human body communication. *IEEE Transac. Biomed. Eng.* 63, 1887–1894. doi: 10.1109/TBME.2015.2504998

Wang, X., Liu, S., Zhu, M., Wang, X., Liu, Z., Jiang, Y., et al. (2019). "Performance of Flexible Non-contact Electrodes in Bioelectrical Signal Measurements," in *In Proceedings of the 2019 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, (Piscataway: IEEE), 175–179.

Wu, H.-T., Chan, Y.-H., Lin, Y.-T., and Yeh, Y.-H. (2014). Using synchrosqueezing transform to discover breathing dynamics from ECG signals. *Appl. Comput. Harmo. Anal.* 36, 354–359.

Yi, Y., Yu, L., Tian, Z., Song, Y., Shao, Y., Gao, L., et al. (2018). Biotemplated synthesis of transition metal nitride architectures for flexible printed circuits and wearable energy storages. *Adv. Funct. Mater.* 28:1805510.

Yu, P., Lu, D., Jin-Zhao, L., Zhang-Yong, L., Qian-Neng, Z., Guo-Quan, L., et al. (2014). A method of removing baseline drift in ECG signal based on morphological filtering. *Acta Phys. Sinica* 63:098701.

Zhang, H., Pei, W., Chen, Y., Guo, X., Wu, X., Yang, X., et al. (2015). A motion interference-insensitive flexible dry electrode. *IEEE Transac. Biomed. Eng.* 63, 1136–1144. doi: 10.1109/TBME.2015.2485269

Zhang, X.-S., Zhu, Y.-S., and Zhang, X.-J. (1997). New approach to studies on ECG dynamics: extraction and analyses of QRS complex irregularity time series. *Med. Biol. Eng. Comput.* 35, 467–473. doi: 10.1007/BF02525525

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 | Frontiers in Neuroscience

ORIGINAL RESEARCH
published: 22 July 2022
doi: 10.3389/fnins.2022.941594

Towards Evaluating Pitch-Related Phonation Function in Speech Communication Using High-Density Surface Electromyography

Mingxing Zhu[1,2†], *Xin Wang*[2,3,4†], *Hanjie Deng*[5], *Yuchao He*[2,4], *Haoshi Zhang*[2,3,4], *Zhenzhen Liu*[6], *Shixiong Chen*[2,4*], *Mingjiang Wang*[1*] *and Guanglin Li*[2,4*]

[1] School of Electronic and Information Engineering, Harbin Institute of Technology, Shenzhen, China, [2] CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, [3] Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen, China, [4] Guangdong-Hong Kong-Macao Joint Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, [5] School of Instrument Science and Engineering, Southeast University, Nanjing, China, [6] Surgery Division, Epilepsy Center, Shenzhen Children's Hospital, Shenzhen, China

Edited by:
Jian Huang,
Huazhong University of Science and Technology, China

Reviewed by:
Wanqing Wu,
Sun Yat-sen University, China
Yi He,
Sun Yat-sen University, China

***Correspondence:**
Shixiong Chen
sx.chen@siat.ac.cn
Mingjiang Wang
mjwang@hit.edu.cn
Guanglin Li
gl.li@siat.ac.cn

† *These authors have contributed equally to this work*

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

Received: 11 May 2022
Accepted: 17 June 2022
Published: 22 July 2022

Citation:
Zhu M, Wang X, Deng H, He Y, Zhang H, Liu Z, Chen S, Wang M and Li G (2022) Towards Evaluating Pitch-Related Phonation Function in Speech Communication Using High-Density Surface Electromyography. Front. Neurosci. 16:941594. doi: 10.3389/fnins.2022.941594

Pitch, as a sensation of the sound frequency, is a crucial attribute toward constructing a natural voice for communication. Producing intelligible sounds with normal pitches depend on substantive interdependencies among facial and neck muscles. Clarifying the interrelations between the pitches and the corresponding muscular activities would be helpful for evaluating the pitch-related phonating functions, which would play a significant role both in training pronunciation and in assessing dysphonia. In this study, the speech signals and the high-density surface electromyography (HD sEMG) signals were synchronously acquired when phonating [aː], [iː], and [əː] vowels with increasing pitches, respectively. The HD sEMG energy maps were constructed based on the root mean square values to visualize spatiotemporal characteristics of facial and neck muscle activities. Normalized median frequency (nMF) and root-mean square (nRMS) were correspondingly extracted from the speech and sEMG recordings to quantitatively investigate the correlations between sound frequencies and myoelectric characteristics. The results showed that the frame-wise energy maps built from sEMG recordings presented that the muscle contraction strength increased monotonously across pitch-rising, with left-right symmetrical distribution for the face/neck. Furthermore, the nRMS increased at a similar rate to the nMF when there were rising pitches, and the two parameters had a significant correlation across different vowel tasks [(aː) (0.88 ± 0.04), (iː) (0.89 ± 0.04), and (əː) (0.87 ± 0.05)]. These findings suggested the possibility of utilizing muscle contraction patterns as a reference for evaluating pitch-related phonation functions. The proposed method could open a new window for developing a clinical approach for assessing the muscular functions of dysphonia.

Keywords: phonation function, pitches, speech communication, high-density, surface electromyogram

INTRODUCTION

Phonation is an essential physiological process *via* which natural voices are produced to aid human communication during daily life activities (Alku, 2011). Fundamentally, a normal voice is characterized by loudness, pitch, and quality, among which pitch is considered the most significant contributor toward constructing intelligible sounds (Kryshtopava et al., 2017). A wide range of normal pitches can enhance vocalization with the rich capability to express one's interpersonal emotions for social communication (Kim et al., 2009; Craig et al., 2015). Pitches used to describe the sound on a scale from low to high are often characterized and quantified in terms of frequencies (in cycles per second or hertz) (Rinaldi et al., 2016). For instance, to deliver a speech or sing a song with high intelligibility, the produced sound should consist of pitches accompanied by harmonic overtones that are multiples of the fundamental frequency (Bishop and Keating, 2012). However, an abnormal pitch resulting in a sound/vocalization that is too high, too low, unstable, monotonous, and unpleasant to listen would lead to phonating difficulty, often known as dysphonia (Jani and Gore, 2014; Knuijt et al., 2014; Sommerville et al., 2017). Also, dysphonia remains a major health issue, especially, for individuals whose jobs require high speech intelligibility such as teachers, singers, lawyers, tour guides, and salespeople, among others (Rosen and Murry, 2000; Cutiva et al., 2013; Martins et al., 2014). Therefore, early examinations and diagnoses of the pitch-related functions of dysphonia could be of great importance for these high-risk populations.

The evaluation of the normality of pitch-related phonation function plays a significant role in diagnosing dysphonia. Currently, most of the methods for assessing the pitch-related phonation functions are based on the analysis of speech signals (Gerratt et al., 2016; Brajot and Lawrence, 2018; Murtola et al., 2019). However, the speech signals are easily affected by environmental acoustic noises and human-introduced interferences (Dennis et al., 2010; Balata et al., 2015). It is difficult to collect speech signals with high quality in a noisy environment, and therefore, the recorded noise-contaminated speech signals were unreliable for assessing pitch-related phonation functions. Moreover, speech signals only contain one-dimension acoustic information about speech, and it is usually insufficient to evaluate the physiological process of phonation activities, the analysis of which requires the dynamic information of the related muscular activities.

Notably, phonation is a complex process controlled by multiple articulatory muscles (Chhetri and Neubauer, 2015). The construction of a normal pitch relies on the oscillation rate of the vocal cords controlled by the contraction patterns of the inherent and extrinsic articulation of the facial and neck muscles associated with phonation (Horáček et al., 2016). The strength, contraction rate, and coordination of the facial and neck muscles are essential factors in the construction of normal pitches during phonation (Macdonald et al., 2012). Therefore, examining the muscular activities during pitch-related phonation tasks is crucial for evaluating the phonating function. It is noteworthy that surface electromyography (sEMG) is an important technique for detecting, recording, and interpreting the electrophysiological characteristics of muscular activities (Tang et al., 2018). Meanwhile, the sEMG approach is non-invasive, safe, easy to operate, and cost-effective when compared with other methods (Naik et al., 2015; Strazzulla et al., 2016), making it widely utilized for assessing the muscular function associated with phonation activities (Pettersen et al., 2005; Van Houtte et al., 2013; Khoddami et al., 2017; Kaneko et al., 2018; Xu et al., 2018). For instance, four channels of sEMG signals recorded while the subjects were phonating a set of vowels at an increasing pitch were utilized to study the electrical activities of scalenus, sternocleidomastoideus, and upper trapezius muscles (Pettersen et al., 2005). In another study, time-domain features of sEMG signals from two channels were used to assess the functions of the cricothyroid and thyrohyoid muscles in patients with dysphonia (Khoddami et al., 2017).

Up to date, most of the previous studies mainly focused on assessing pitch-related phonation functions individually using either speech signals (Gerratt et al., 2016; Brajot and Lawrence, 2018; Murtola et al., 2019) or sEMG recordings (Pettersen et al., 2005; Van Houtte et al., 2013; Khoddami et al., 2017; Kaneko et al., 2018; Xu et al., 2018), whereas the quantitative interrelationship between the pitches and the corresponding muscular activities during the pitch-related phonation remains unclear. It is noteworthy that the interrelations between the pitches and the corresponding muscular activities are an essential prerequisite for exploring the electrophysiological mechanisms of pitch-related phonation functions, and such mechanisms play an important role in pronunciation training and dysphonia diagnosing. Therefore, the synchronous analysis of both speech signals and sEMG recordings is necessary to clarify the interrelations between the pitches and muscular activities.

Furthermore, since the phonation process involves a large group of small facial and neck muscles that span a relatively large area (Dewan et al., 2017), dynamically evaluating the electrophysiological spatiotemporal properties of the entire group of coordinated facial and neck muscles could be helpful for the analysis of the pitch-related phonation functions. However, most of the existing studies utilized only a few surface electrodes (typical 2–4 channels) with limited coverage, and they rarely provided adequate neuromuscular information required for consistently accurate diagnoses (Pettersen et al., 2005; Khoddami et al., 2017). A small number of sEMG electrodes might miss important muscles and major electrical activities that would be essential for the dynamic assessment of the entire phonation process. Therefore, a large number of sEMG electrodes would be necessary to cover the whole group of facial and neck muscles in enough density so that the spatiotemporal properties and coordination activities of all the muscles could be studied systemically.

High-density sEMG (HD sEMG) is a non-invasive technique to measure sEMG signals with a two-dimension array of closely spaced electrodes, and it could provide comprehensive information for dynamic evaluation of the spatiotemporal characteristics of muscle groups (Johns et al., 2016; Zhu et al., 2017b, 2018b; Chen et al., 2018; Glaser and Holobar, 2018). When applied in the phonation study, the HD sEMG technique could be beneficial for collecting sufficiently relevant information to evaluate the properties of the facial and neck muscles across the entire phonation process. In this regard, a series of sequential energy maps constructed from HD sEMG recordings could be utilized to evaluate the muscular activities of the frontal neck while phonating vowel [aː] in our pilot studies (Zhu et al., 2017a, 2018a). Afterward, Bracken et al. utilized a 20-channel array of HD sEMG signals located around the anterior neck to evaluate the muscular activities when phonating in three manners (rest, low, and high pitches) (Bracken et al., 2019). Nevertheless, these studies only focused on myoelectric characteristics of the neck muscles without simultaneously considering the properties of speech signals that are commonly applied in clinical settings, and therefore, the interrelations between the speech signals and the sEMG recordings during pitch-related phonation have rarely been investigated.

In this study, we investigated the possibility of utilizing the acquired dual-signal (speech and HD sEMG signals) for systematically exploring the interrelations between the pitches (sound frequencies) and the corresponding muscular activities associated with pitch-related phonation. This purpose had been achieved by dynamically visualizing the electrophysiological spatiotemporal properties of the facial/neck muscles and quantitatively analyzing the correlations between sound frequencies and myoelectric characteristics when phonating with increasing pitches across different vowel tasks. The proposed method might open a new window for developing a clinically relevant approach for training the pronunciation and diagnosing the dysphonia.

METHODS

Subjects

Toward reducing the complexity of the subjects, a total of 14 healthy male volunteers (mean age = 24.7 ± 1.5 years) participated in this study involving a set of systematically designed phonation tasks. Before the experiments, all the subjects were enquired to ensure that they had no history of phonation difficulties, and a pure-tone audiometry test was conducted to ensure that they had no hearing problems. Then, the objective and experimental procedures of the experiments were clearly explained before the data collection session. All the subjects gave written informed consent and provided permission to publish their photographs/data for scientific and educational purposes. The protocol of this study was approved by the Institutional Review Board of Shenzhen Institutes of Advanced Technology (#IRB ID: SIAT-IRB-170815-H0178).

Experimental Procedures

In this study, the speech signals and the HD sEMG recordings were simultaneously acquired from all the recruited subjects when a subject phonated the back vowels [aː], front vowels [iː], and central vowels [əː] with a continuously increased pitch. The speech signals were recorded at a sampling rate of 44,100 Hz *via* a headset-attached microphone placed above the first row of sEMG electrodes on the right side of the face, as seen in **Figure 1A**. Meanwhile, the HD sEMG signals were recorded by using a multichannel EMG recording system (TMSI, REFA, the Netherlands) at a sampling rate of 2,048 Hz. During the experiment, we implemented a hardware circuitry that could trigger the start of the data acquisition of both the speech and HD sEMG to achieve the synchronization of the two types of signals. By using double-sided small-sized tapes, a total of 120 channels of surface electrodes were evenly placed on the skin surface of the face and the neck, both with left/right symmetry (**Figure 1B**). The double-sided tapes were medical grade and had very strong adhesive strength to stick the electrode on the skin to avoid the interferences introduced by the electrode wire movements or lip motion. Before placing each electrode, the skin of the subject was carefully prepared with an abrasive skin preparation gel and then cleaned with alcohol cotton to remove any sweat and oil that could possibly reduce the adhesion of the double-sided tapes.

Each surface electrode is ∼10 mm in diameter, and an inter-electrode distance of about 15 mm was ensured between two neighboring electrode centers to cover the phonation-related muscles in intensive densities. In this regard, a two-dimension (2D) array made up of 80 electrodes in a 5 × 16 grid was placed on the suprahyoid and infrahyoid muscles located in the front neck regions of each subject, as seen in **Figure 1B**. The electrodes were identified and labeled by 5 rows (E–I) and 16 columns (1–16). The remaining 40 electrodes were symmetrically placed over the facial muscles (including the masseter, the buccinators, the orbicularis oris, and the zygomatic minor muscles), with a 4×5 grid on both the left and right sides. The 2D facial electrodes were identified and labeled by 4 rows (A–D) and 10 columns (1–10), as seen in **Figure 1B**. The extra space in the middle between the left and right neck electrode arrays was to avoid the laryngeal prominence of the subjects, especially in male subjects. Afterward, a fabric electrode was placed on the left wrist of the subject to serve as the reference electrode.

The phonation experiments were individually carried out on all the recruited subjects in an electromagnetic-shielded and soundproofed room that allowed the collection of high-quality electric and acoustic signals. Three different phonating experiments were adopted in the current study to assess the muscle activation patterns and to explore the interrelations between the speech and HD sEMG signals. First, the subjects were asked to maintain a quiet state without phonating or moving their body parts for around 30 s so that the baseline for the speech and the HD sEMG signals could be determined. Then, the subjects were asked to pronounce [aː], [iː], and [əː] vowels, respectively, in a sequential order with a monotonously rising pitch (from low to high) for three successive repetitions per trial. In each repetition, the target vowel was phonated for ∼4 s and followed by a rest period of 6 s that comprised 3 s before the phonation and 3 s after the phonation to prevent the subjects from experiencing either muscle or mental fatigue that may compromise the fidelity of the signals. During the phonation, the speech signals and the HD sEMG recordings were synchronously acquired from all the subjects.

Analysis of the Speech Signals

The spectrogram of the speech signals during the phonation task was derived by using the short-time Fourier transform (STFT)

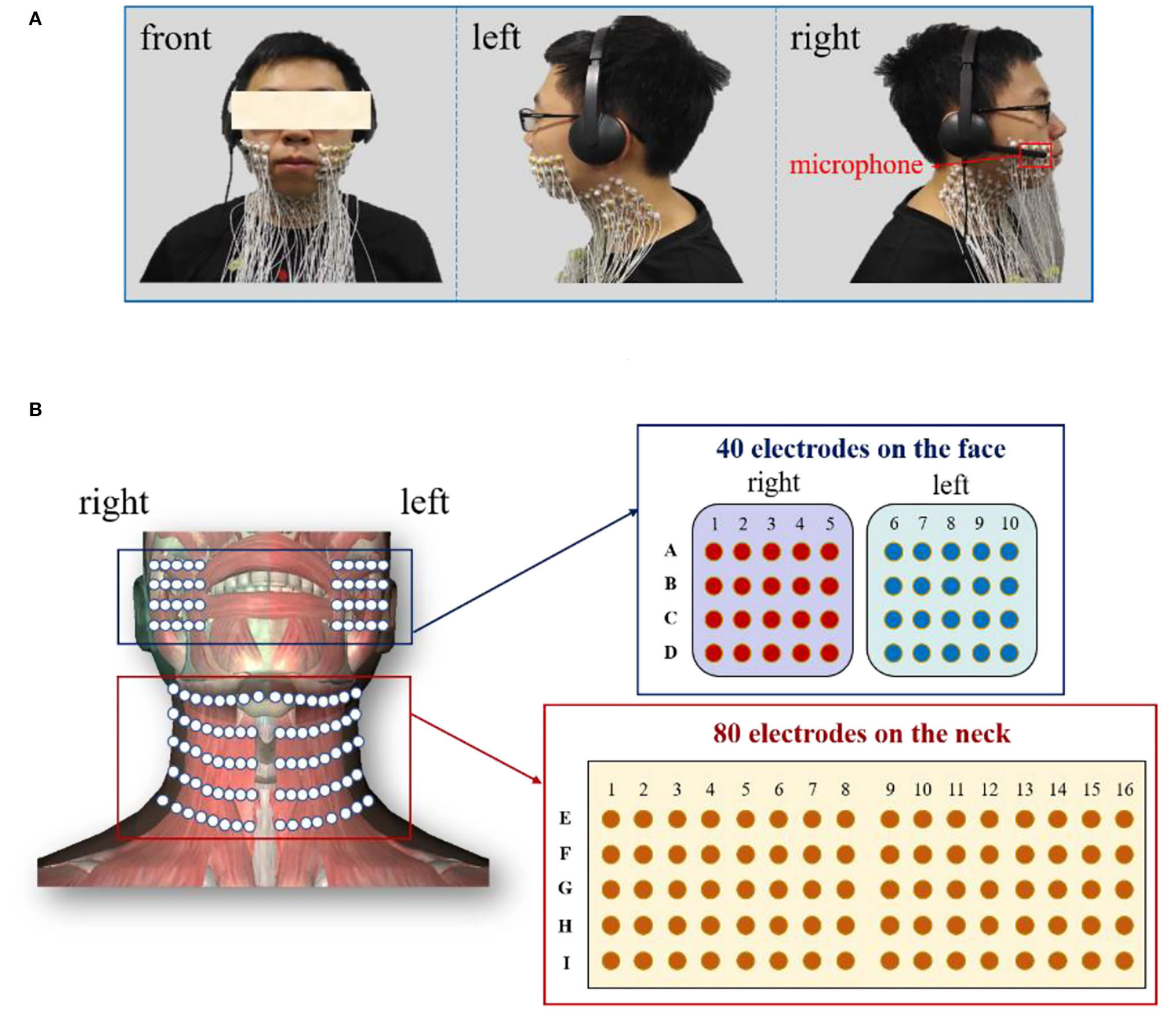


FIGURE 1 | (A) Experimental setup to record the speech and HD sEMG signals simultaneously; **(B)** Placement of the high-density surface electrodes on the facial and neck muscles.

method to perform the time-frequency analysis. The duration of the STFT analysis window was 20 ms, and the overlap width between two neighboring windows was 10 ms. The formula of the STFT method implemented in this study could be expressed as follows:

$$\mathrm{STFT}\left\{x\left[n\right]\right\}\left(m,\omega\right)=X\left(m,\omega\right)=\sum_{n=-\infty}^{\infty}x\left[n\right]\omega\left[n-m\right]e^{-j\omega n} \qquad (1)$$

where x[n] is the segmental speech signal at time frame n and ω[n-m] is a Hamming window with the same duration as x[n].

To further illustrate the frequency domain features, the speech signals obtained during the phonation tasks were segmented into a series of frames with a duration of 250 ms, and the amplitude spectrum of each frame was calculated accordingly. Then, frame-by-frame spectral curves could be obtained to demonstrate the frequency-domain characteristics of the speech signals when the subject was phonating vowels in an increasing pitch. Afterward, the median frequency (MF) of the speech signals from all frames was computed. MF is defined as the frequency at which the spectrum is divided into two regions with equal amplitude; in other words, MF is the frequency at half of the power spectral density of the signals (Phinyomark et al., 2012). It can be expressed as given in Equation (2). Finally, the MF values were normalized by employing the min-max normalization method given in Equation (3).

$$\sum_{j=1}^{MF} p_j = \sum_{j=MF}^{M} P_j = \frac{1}{2}\sum_{j=1}^{M} P_j \tag{2}$$

where P_j is the EMG power spectrum at frequency bin j and M is the length of the frequency bin.

$$nMF(i) = \frac{MF(i) - \min(MF)}{\max(MF) - \min(MF)} \tag{3}$$

where nMF(i) is the normalized MF value of speech signals, *MF(i)* is the MF value in analysis window *i*, min(MF) is the minimum MF value, and max(MF) is the maximum MF value during the entire phonation.

Analysis of the HD sEMG Signals

A set of digital filters were used to enhance the quality of the HD sEMG signals obtained during the different phonation tasks. In this regard, we first used a third order Butterworth band-pass filter from 10 to 500 Hz to reduce baseline fluctuations. Then, a Butterworth band-stop filter was also applied to attenuate the power line interferences at 50 Hz and its harmonics. After the pre-processing, the filtered sEMG signal of each channel was sliced into a series of sequential 250 ms analysis windows, and the root-mean square (RMS) was computed for each analysis window to obtain the average energy distribution of the muscular activities as follows:

$$\mathrm{RMS}\,\{sEMG\,[n]\} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} sEMG^2\,[i]} \tag{4}$$

where RMS{*sEMG*[*n*]} is the RMS value of sEMG signals for each analysis window, *sEMG*[*i*] is the *i*th sample in the analysis window, and *n* is the total sample number of the analysis window.

Then, the frame-by-frame RMS values of all the 120 channels were computed to investigate the dynamic activities of the facial and neck muscles during the phonation with continuously increasing pitches. Besides, RMS values of the same time frame from different neighboring channels were joined together to form a 2D array according to the electrode positions so that a 2D energy map could be obtained to present the energy distribution of the corresponding region. In this way, an energy map with a size of 4 × 5 could be obtained for the left and right faces, respectively. A 5 × 16 energy map could also be generated for the neck region. Then, the energy maps of all the time frames were globally normalized for the face and neck region, respectively, so that the sequential energy maps in a frame-wise manner could help to demonstrate the spatial and temporal changes in energy and muscular activities during different phonation stages.

$$nRMS(i) = \frac{RMS(i,j) - \min[RMS(i)]}{\max[RMS(i)] - \min[RMS(i)]} \tag{5}$$

where nRMS(i) is the normalized RMS value of sEMG signals in channel *i*, *RMS*(*i*, *j*) is the RMS value of channel *i* in analysis window *j*, min[RMS(i)] is the minimum RMS value of channel *i*, and max[*RMS*(*i*)] is the maximum RMS value of channel *i*.

Correlation Analysis Between Speech and sEMG Signals

The nMF of the speech signals, as well as nRMS values of the HD sEMG signals per analysis window, was obtained and compared to investigate the correlations between the speech signals and muscular activities. The entire phonation process with a duration of 10 s was segmented into a series of analysis windows with a length of 250 ms, resulting in a total of 40 nMF and 40 nRMS values. Then, the correlation coefficients between the two groups of nMF and nRMS were calculated to analyze the interrelations between the speech and sEMG signals.

RESULTS

Features of the Speech Signals With an Increasing Pitch

In this study, the time-domain and time-frequency characteristics of the speech signals obtained during the phonation process of [a:] vowel using low to high pitch were analyzed as follows. The waveform characteristic of the speech signals in the time-domain is seen in **Figure 2A**. From this figure, the speech signals were observed to immediately rise to ∼76 dB SPL (calibrated through a professional sound level calibrator, Model AWA6021A, Gester Instruments Co., China) at the commencement of the phonation, and about 4 s later, the volume of the signal fluctuated around 80 dB SPL, thus exhibiting a relatively stable phenomenon when the pitch increased continuously from low to high. The speech signals instantaneously returned to the baseline at the completion of the phonation task.

Meanwhile, the time-frequency characteristics of the speech signals along the time axis are presented in **Figure 2B**. According to the analysis herein, the time-frequency distribution during phonation revealed that the high intensity of the speech signals was obtained from low-frequency components, especially between 0 and 8 KHz, which covered the entire phonating activities. Additionally, the frequency values of the speech signals were observed to increase steadily with a corresponding rise in pitch when [a:] vowel was phonated. Also, the highest frequency components appeared at the end of the phonation, where the pitch also got to its peak.

A series of waveforms across the phonation process of [a:] vowel was constructed to characterize the amplitude spectrum of the speech signals. In this regard, the speech signals were segmented into 40 continuous windows of a length of 250 ms, otherwise known as frames from which the amplitude spectrum waveforms were later obtained and presented in a successive manner, as seen in **Figure 3**. Each spectrum showed the amplitude of each individual frequency within the corresponding time frame of speech to visualize the dynamic variation in the frequency domain. The amplitude spectrum of the frames (F1–F13) that correspond to the speech signals obtained before

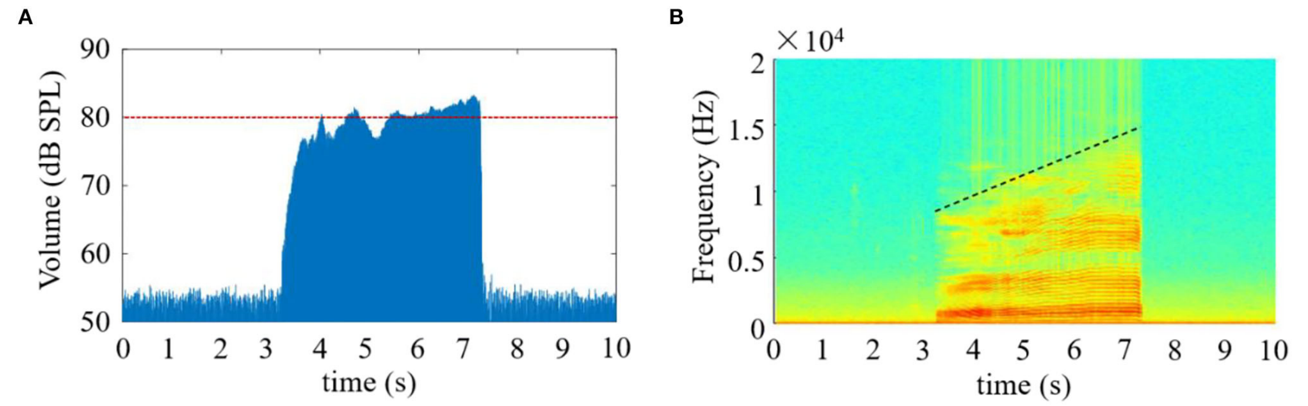


FIGURE 2 | The amplitude and spectrogram of the speech signals when phonating the vowel [a:] with continuously increasing pitch. **(A)** The amplitude of the speech signal as a function of time; **(B)** The time-frequency spectrogram of the speech signal.

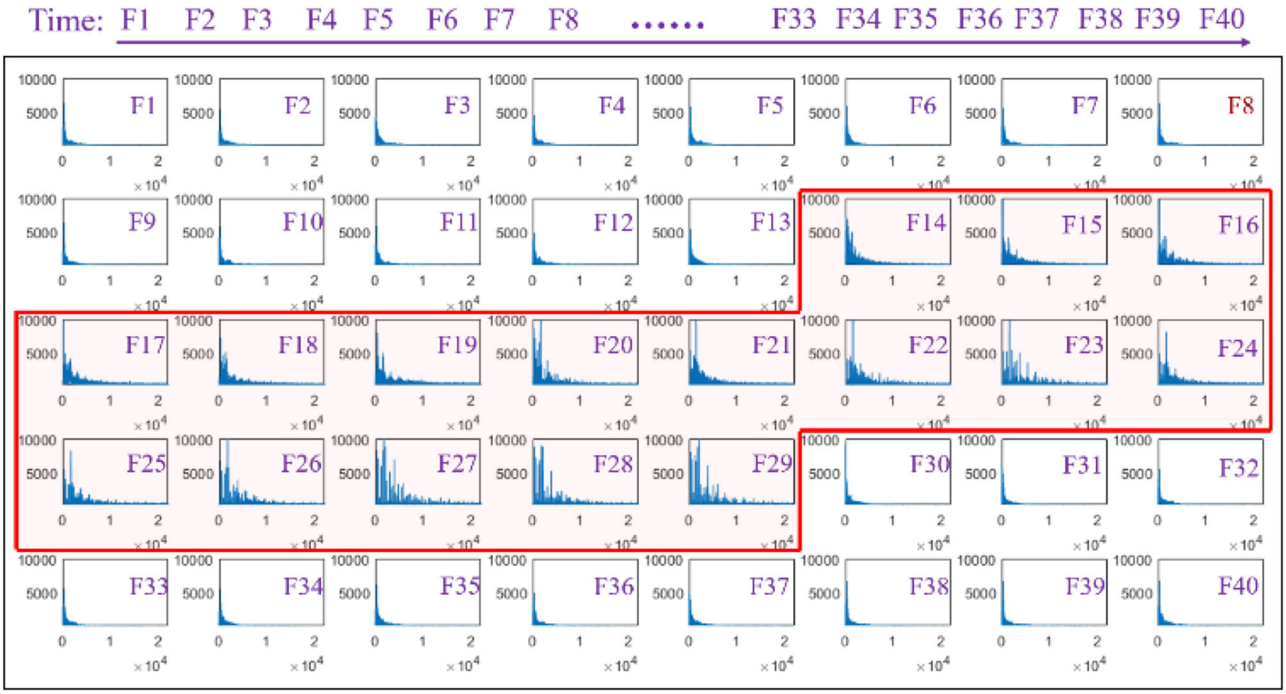


FIGURE 3 | Amplitude spectra of the speech signal with an increasing pitch from a series of analysis windows (F1–F40).

phonating the vowel [(a:)] was observed to have a low-frequency band with low intensities. Meanwhile, the amplitude spectra were observed to vary considerably between the 14th and 29th frames when the subjects began phonating at a low pitch and steadily increased to a high pitch. Specifically, at the beginning of the phonation task represented in frames F14–F16, the frequency band of the speech signals increased slightly to values below 10 KHz with a corresponding increase in the amplitude of the frequency. With a further increase in the pitches, the frequency band of the signals became even wider from frames F17–F29 with a band that was ∼20 KHz. Meanwhile, the amplitude of the frequency also increased steadily from F17–F29 with a corresponding increase in pitches. Subsequently, the frequency band and the amplitude of the speech signals suddenly dropped when the subjects stopped phonating.

Temporal Waveforms of the HD sEMG Signals Across the Phonation Process of [a:] Vowel With a Rise in Pitch

The temporal waveforms of HD sEMG signals are displayed in **Figure 4** for showing the characteristics of facial and neck

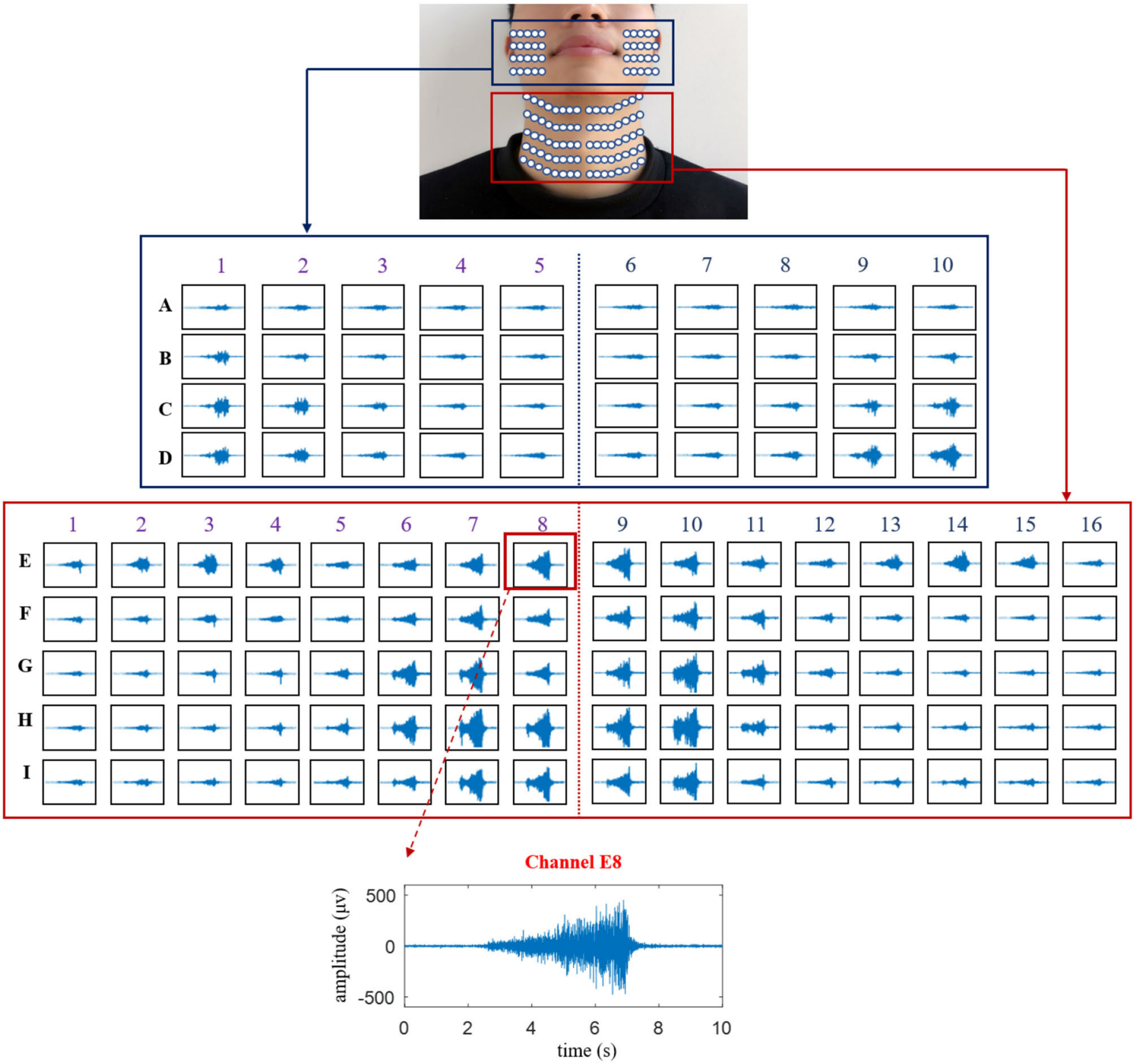


FIGURE 4 | The temporal waveforms of the HD sEMG signals recorded from all the 120 surface electrodes on the facial and neck muscles during the phonation with rising pitch.

muscles, in which the waveform of a typical channel (E8) was presented to exhibit the variation of the sEMG amplitudes. At the onset of the phonation task, the temporal waveforms of the HD sEMG signals associated with the facial and neck muscles had relatively lower intensity, and as the pitch increases from low to high, the amplitude of the waveforms also increased correspondingly across all the channels. Meanwhile, at the offset of the phonation task, the HD sEMG waveforms were observed to have a sudden drop in amplitude to the baseline.

As seen in **Figure 4**, the HD sEMG waveforms were constructed for 120 channels that were divided into two parts: the first denotes the facial region (in the blue rectangle) consisting of 40 channels (from columns 1 to 10 and rows A to D) and the other part indicates the neck region (in red rectangle) consisting of 80 channels (from columns 1 to 16 and rows E to I). On the facial region (row A to D), the amplitudes of the sEMG waveforms obtained from the electrodes located around the mouth (from columns 3 to 8) were relatively smaller than those obtained from the signals on the outer part of the face (right side: columns 1–2 and left side: columns 9–10) during the entire phonating process. On the contrary, as shown in the neck region (rows E–I), the amplitudes of the sEMG signals turned out to be larger at the center of the neck area (from columns 6 to 11) than those at the edge of the neck (right

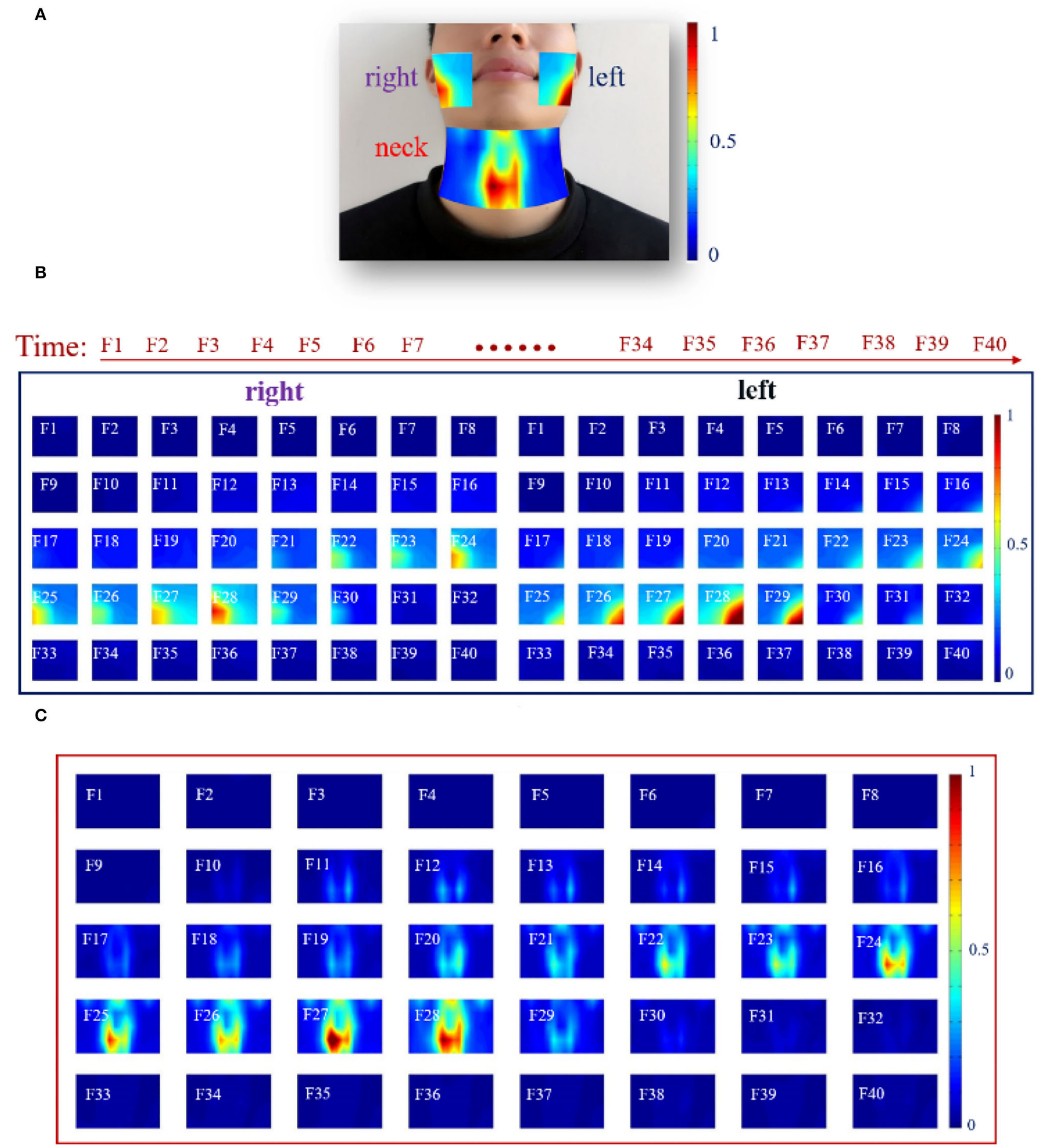


FIGURE 5 | The HD sEMG energy maps of the facial and neck muscles by using 40 analysis windows with a length of 250 ms during phonating vowel [aː] with pitch increase: **(A)** A typical frame of HD sEMG energy map visualizing the facial and neck muscle activities; **(B)** HD energy maps of the facial muscles; **(C)** HD energy maps of the neck muscles.

side: columns 1–5 and left side: columns 12–16). Meanwhile, the sEMG signals on the first row of the neck region (E1–E16) also showed high intensities when phonating with an increasing pitch. Additionally, the symmetrical distributions of the HD sEMG waveforms were observed between the left and right sides for both facial and neck regions.

Dynamic HD sEMG Energy Maps When Phonating [a:] Vowel With Increasing Pitch

In this section, dynamic HD energy maps were constructed to analyze the characteristics of the facial and neck muscles during the entire phonation process. For visualizing the dynamic changes in muscular activities associated with facial and neck muscles during rising pitches, the earlier described RMS feature was computed from a series of analysis windows that resulted in feature matrices utilized for generating the HD sEMG energy maps presented in **Figure 5**. Note that different intensities of myoelectric activities are represented by the color gradient of the energy maps, in which the color toward red signified high intensity and that toward blue denoted low intensity.

First, a typical dynamic HD sEMG energy map of the facial and neck muscles from an analysis window during the phonation process is displayed in **Figure 5A**. It could be seen that the map in the neck region had higher intensity, especially for muscles that appeared around the center of the neck. Thus, the muscles in the active neck region were distinctively separated into two parts symmetrically, in which the right side had slightly higher energy intensity than the left region of the neck. Meanwhile, low-intensity energy was observed on the top of the map that is symmetrical between the right and left sides. In addition, a highly symmetrical energy distribution was equally observed between the right and left sides. The energy map on the right side of the facial muscles had high intensity on the edge of the right face region, while the map on the left side of the facial muscles reflects high intensity on the left side of the facial muscles (**Figure 5A**).

Subsequently, a further analysis that involved the entire phonation process was done to observe if similar results could be obtained from a single channel described earlier. In this regard, the entire phonation task was segmented into 40 analysis windows of a length of 250 ms each, and a sequence of HD sEMG energy maps was constructed to reflect the dynamism of the myoelectric activities associated with the facial and neck muscles. Hence, a total of 40 frames of HD sEMG energy maps representing myoelectric activities of the facial muscles are shown in **Figure 5B**, while 40 frames of HD sEMG energy maps of the neck muscles are presented in **Figure 5C**.

At the onset of the phonation task (F1F10), the subjects were asked to assume a quiet state without moving. Hence, no activity could be seen for both the facial (**Figure 5B**) and neck (**Figure 5C**) muscles. Meanwhile, when the subjects began phonating, the HD sEMG energy maps became different between the facial and neck muscles, as could be seen from F11 to F30 in **Figures 5B,C**.

As seen in **Figure 5B**, the facial muscles were kept in unconspicuous activities on both the right and left sides from F11 to F20 even after the subject had already begun to increase the pitch. It was noticed that there was some delay in the facial muscle activation during the phonating process. Then, from F22 to F28, the intensity of the energy maps started constantly increasing to reach the maximum value on F28, and the high-intensity areas of the maps on the right and left sides were concentrated on the edge of the facial muscles. The active regions on the left side of the facial muscles appeared at the bottom left of the maps, while that on the right side of the facial muscles performed at the bottom right of the maps. Additionally, it was found that the HD sEMG energy maps of the facial muscles were symmetrically distributed on both the left and right sides, besides there were some differences between the two sides, in which the intensities of the left facial muscles were a little higher than that of the right facial muscles. After that, at the end of the phonating process, the myoelectric activities of the facial and neck muscles were diminished in a short time (from F29 to F30) when the subject prepared to stop pronouncing. Finally, after the subject went back to quiet, the high-intensity area of the energy maps disappeared from F31 to F40.

Then, it was observed that the muscular activities of the neck muscles (**Figure 5C**) were activated before the facial muscles (**Figure 5B**). The high intensity of the energy maps on the neck muscles started to emerge from the center of the map at the 11th frame in low intensity, while that of the facial muscles was activated at the 21st frame (**Figure 5C**). Afterward, the energy maps were maintaining a stable intensity along with the following five frames from F12 to F17. Later, from F18 to F28, the energy maps showed that the intensities in the center region of the neck muscles were gradually increased and spread upward to the upper locations when the subject was phonating vowel [a:] with continuous rising pitches. During the whole phonating process, the muscle activations were concentrated on the center of the neck muscles and split up into two parts by the centerline. Meanwhile, the energy maps clearly presented approximately symmetric activity regions on the left and right sides of the neck muscle. Also, the total intensity of myoelectric activities on both the right and left sides reached a maximum value at the 28th frame. Thereafter, at the end of the phonating process, the intensities of energy maps were diminished in the 29th frame in a short time. Finally, when the subject stopped phonating, the high-intensity area of the energy maps disappeared from F30 to F40 in **Figure 5C**. Besides, similar energy patterns of the dynamic HD sEMG maps on facial and neck muscles could also be observed in the other 14 subjects when they were phonating vowel [a:] with pitch increase.

Interrelations Between Sound Frequencies and Myoelectric Activities During Phonating [a:] Vowel

To quantitatively investigate the interrelations between the pitches of the speech signals and the intensity of the myoelectric activities while phonating vowel [a:] on a rising pitch scale, the nMF of speech signals and the nRMS values of sEMG recordings were compared across the entire phonating process. Then, the correlation coefficient between the nMF and nRMS features was computed across three repetitions of the phonation task, and the obtained result for a representative subject is presented in **Figure 6**.

In **Figure 6A**, the curves in blue showed the normalized nMF of the speech signals, while that in red represented the nRMS of the sEMG signals for a randomly selected channel. It could be observed that the waveforms of the nMF and nRMS features exhibited similar morphology in the first test session (test 1), in which the nRMS values (from sEMG signals) increased in

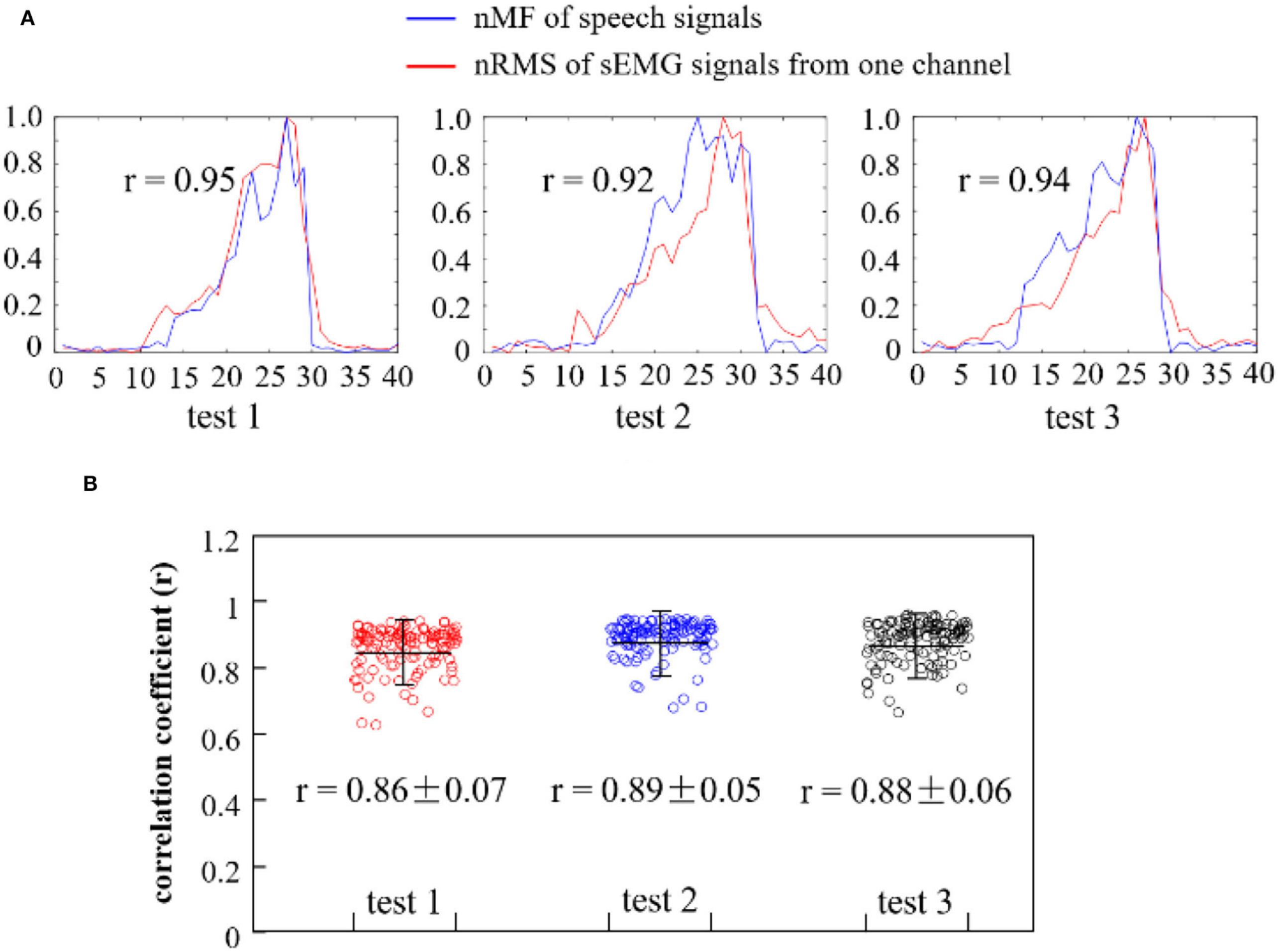


FIGURE 6 | (A) Comparison of the nMF of the speech signals and the nRMS of the sEMG recordings from one channel when phonating vowel [a:] with increasing pitch; **(B)** Correlation coefficients between the nRMS values of 120 channels of sEMG recordings and the nMF of the speech signal from three repeated tests.

a similar rate with the nMF values (from speech signals) at a constantly rising pitch. It could be seen that when the pitch reaches its maximum value, both the nMF and nRMS values rapidly declined in a similar manner until they both hit the baseline. This phenomenon is also observed for the other two test sessions (test 2 and test 3) in **Figure 6A**. The nRMS feature curve appeared before the nMF feature curve at the onset of the phonation task, while the nRMS feature curve disappeared after the nMF feature curve at the offset of the task. Furthermore, a significant correlation was observed between the nMF and the nRMS features ($r = 0.95$ in test 1, $r = 0.92$ in test 2, and $r = 0.94$ in test 3) at a $p < 0.05$ based on Pearson's correlation coefficient analysis. In other words, this analysis indicated that a strong correlation existed between sound frequencies and muscular activities associated with pitch-related phonation.

To further validate the above claim, the correlation coefficient between the nMF of the speech signals and the nRMS across the 120 channels of sEMG recordings in the three repeated tests was computed, and the obtained result is presented in the scatter plots seen in **Figure 6B**. The red, blue, and black scatterplots separately showed the correlation coefficients corresponding to test 1, test 2, and test 3. The correlation coefficients were closely clustered with an average value of 0.86 in test 1, 0.89 in test 2, and 0.88 in test 3 for a representative subject, indicating that there was a significant correlation between the nMF of the speech signals and the nRMS of the sEMG signals across the three repeated tests. Additionally, a few points were seen to be dispersed away from the center of the cluster, which could be a result of either random noises or unintentional body movements that must have occurred during the data acquisition session.

Averaged Dynamic HD sEMG Energy Maps When Phonating [i:] and [ə:] Vowels on a Rising Pitch Scale

The effects of the facial and neck muscle activities across different vowels ([i:] and [ə:]) during the normal phonating processes were also investigated using the HD sEMG energy maps from another subject. Each phonating task that contained 40 frames of energy maps was truncated to 24 frames (from frame 9 to 32). Then, two

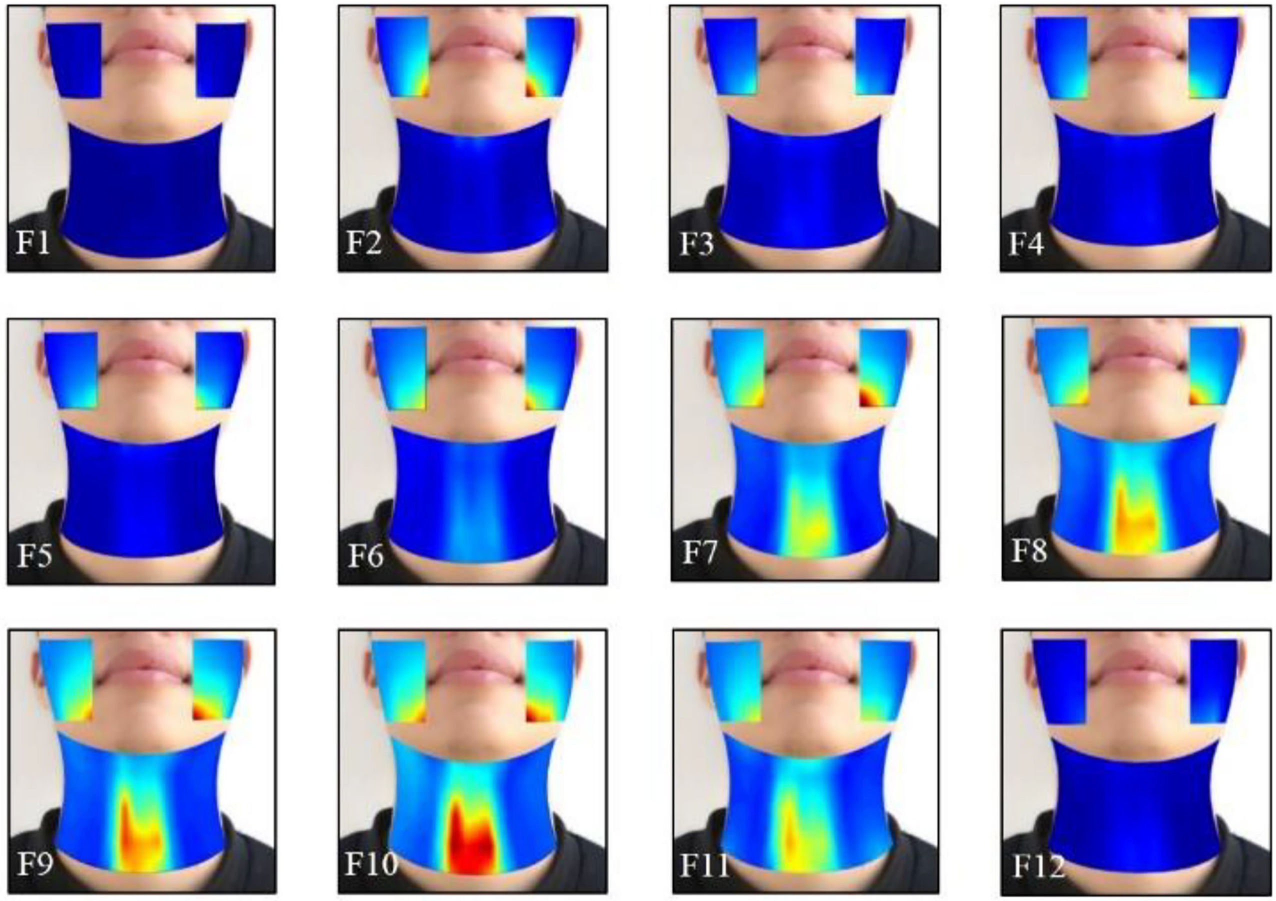


FIGURE 7 | The averaged HD sEMG energy maps of the facial and neck muscles during the phonating vowel [i:] with a pitch increase.

of the adjacent frames from the remaining 24 maps were averaged with non-overlapped epochs to construct 12 consecutive energy maps (F1–F12). It can be seen from **Figures 7**, **8** that the HD sEMG energy maps from the two phonating tasks showed the explicit difference with regard to the duration and strength of the facial and neck muscle contractions.

Figure 7 showed the averaged dynamic sEMG energy maps when one subject was phonating [i:] vowel. When starting phonating, the high-intensity areas were first observed on the sEMG maps of the facial muscles (from F2 to F5), while the activities of the neck muscles were kept at low intensity. There was some delay in the neck muscle activation than the facial muscles during the phonating process. Then, from F6 to F10, the intensity on the energy maps of both the facial and neck muscles started increasing continuously to reach the maximum value on F10. The high-intensity areas of facial muscles were concentrated below the angulus oris on the right and left sides, while the areas were over the center of the neck muscles with high symmetrical distribution on the left and right sides of the neck. Meanwhile, the activated areas of the facial and neck muscles were expanded during increasing pitches (F6–F10). After that, the intensity of the maps was diminished in a short time at frame 11 when the subject prepared to stop phonating. Finally, the high-intensity area of the energy maps disappeared in frame 12 at the end of the phonating tasks.

The HD sEMG energy maps in **Figure 8** represent a sample of the energy distribution of the facial and neck muscles when phonating the vowel [ə:] on a rising pitch scale, which was observed to be consistent in the other 13 subjects. Before the phonation, the maps showed a low intensity from the 1st through the 5th frame. Then, the intensities of the energy maps on the neck muscles were gradually increased from frame 6 to 10, while the maximum intensity was achieved in frame 10. The areas of the high intensity were also gradually expanded with the symmetrical distribution of both the left and right sides. Meanwhile, the high light areas of the map on the facial muscles first appeared at the bottom edge of the maps in low intensity (in F8), where the activation time was delayed than the neck muscles. After that, the intensities on the maps were increased until F10 on both the left and right sides of the face and the highlighted areas were enlarged when increasing pitches. Then, the intensity of the energy maps on the facial areas reduced at F11 and finally disappeared at F12 after the subject stopped phonating (**Figure 8**). The intensities of the neck muscles were higher than that of the neck muscles,

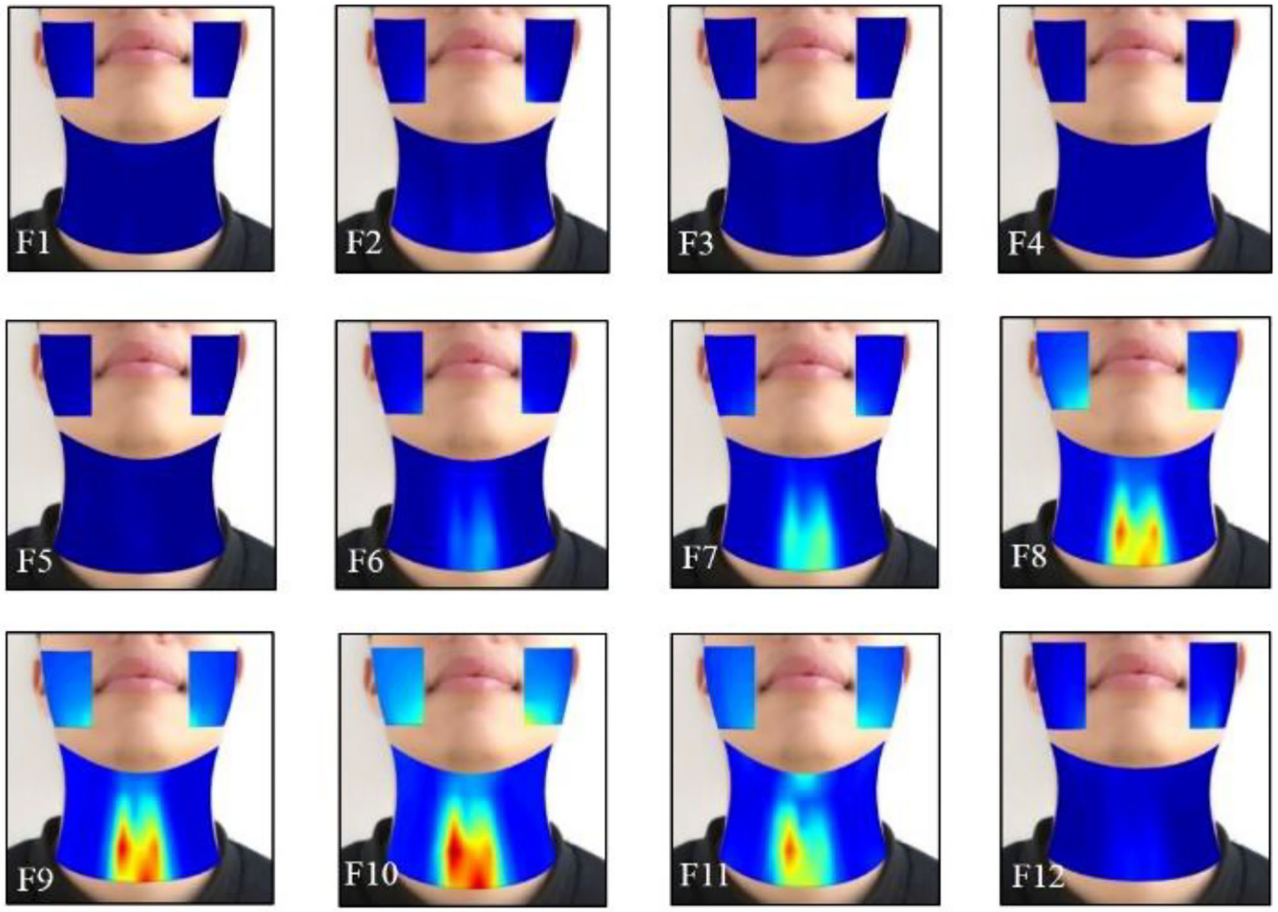


FIGURE 8 | The averaged HD sEMG energy maps of the facial and neck muscles during the phonating vowel [ə:] on a rising pitch scale.

which revealed that the neck muscles were more active than the facial muscles when phonating [ə:].

Correlations Between the nMF of the Speech Signals and the nRMS of the sEMG Recordings Across Different Vowel Tasks

In this section, the correlations between the nMF of the speech signals and the nRMS of the sEMG recordings were investigated for the three phonating tasks including the back [(a:)], front [(i:)], and central [(ə:)] vowels across subjects when the pitch increased from low to high, and the obtained results are shown in **Figure 9** and **Table 1**.

In **Figure 9**, the nRMS values from one channel of sEMG (red lines) and the nMF values (blue lines) from the speech signals for a specific subject were examined and compared across the three different phonating tasks [(a:), (i:), and (ə:) vowels]. From **Figure 9A**, it could be seen that the nMF of the speech signals was gradually rising when the phonating vowel [a:] was with a constantly increasing pitch and suddenly drop to the baseline after the maximum pitch is attained, indicating the end of the phonation task. Meanwhile, the nRMS of the sEMG recordings also exhibited similar characteristics as the nMF at the onset, intermediate, and offset points. However, the curve of the nRMS got initiated before that of the nMF at the onset of the phonation task and as well as hit the baseline after the nMF curve at the offset of the task. In like manner, the phenomenon observed between nMF and nRMS while phonating the [a:] vowel was also observed when phonating [i:] vowel and [ə:] vowel. In summary, significantly high correlations were recorded between the nMF of the speech signals and the nRMS of the sEMG signals during the three phonation tasks: [a:] ($r = 0.95$, $p < 0.05$), [i:] ($r = 0.94$, $p < 0.05$), and [ə:] ($r = 0.91$, $p < 0.05$).

Moreover, to investigate the interrelations between the nMF of the speech signals and the nRMS from the 120 channels of sEMG recordings across different vowel tasks, the correlation coefficient of the two parameters was calculated and displayed as box plots shown in **Figure 9B**. The blue-violet, yellow, and red boxes in **Figure 9B** show the correlation coefficients corresponding to phonating [a:], [i:], and [ə:] vowels, respectively. The correlation coefficients had overall average values of 0.90 ± 0.05 in [a:] vowel task, 0.82 ± 0.08 in [i:] vowel task, and 0.86 ± 0.04 in [ə:] vowel task, which revealed that there was a significant correlation between the nMF of the speech signals and the nRMS of the

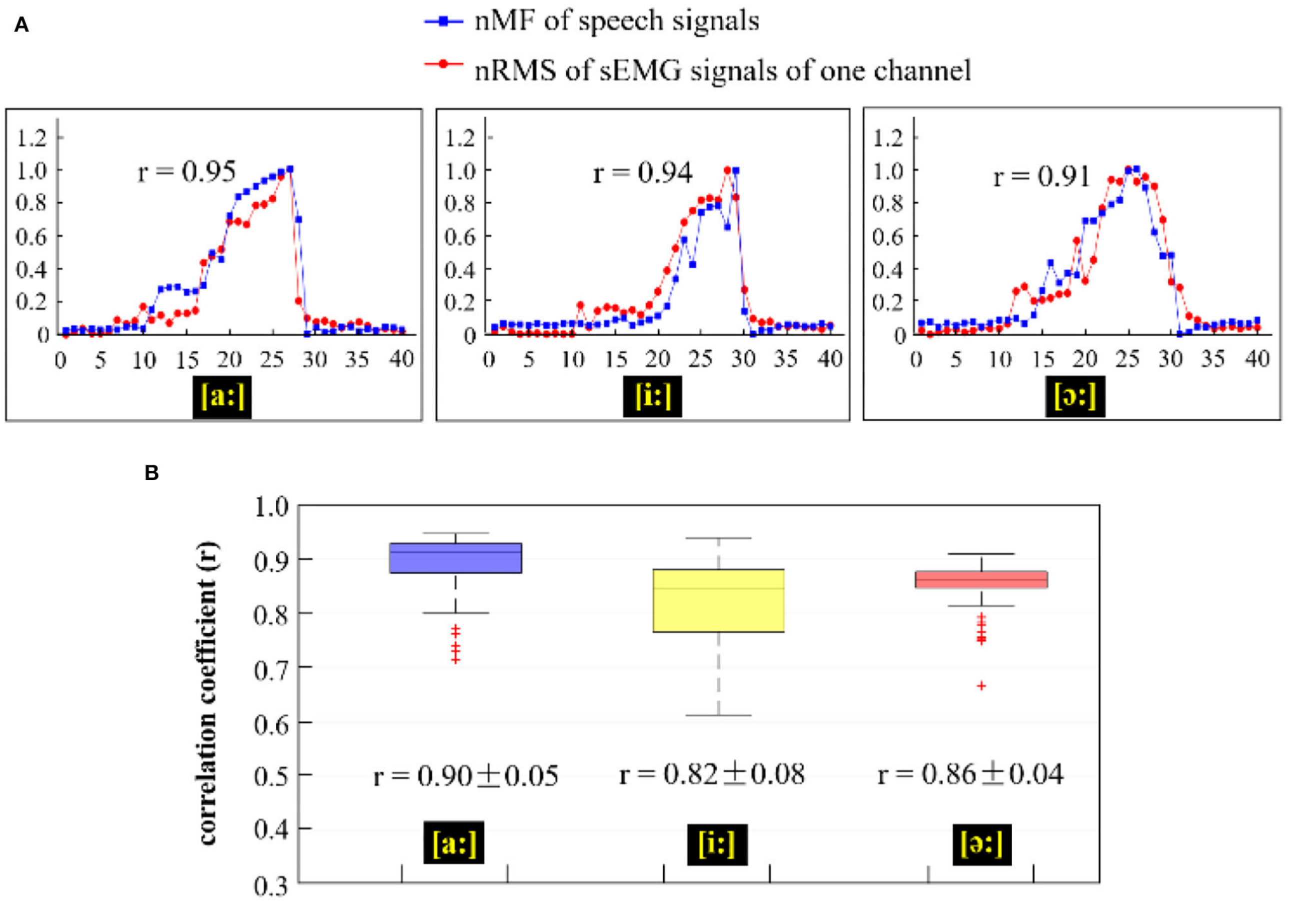


FIGURE 9 | (A) Comparison of the nMF of the speech signals and the nRMS of the sEMG recordings from channel E8 when phonating three different vowels [(a:), (i:), and (ə:)] with increasing pitch; **(B)** Correlation coefficients between the nRMS values of 120 channels of sEMG recordings and the nMF of the speech signal across three different vowels [(a:), (i:), and (ə:)].

TABLE 1 | The average correlation coefficients between the nRMS of sEMG recordings from all channels and the nMF of the speech signals across all recruited subjects (*p*<0.05).

	Sub	1	2	3	4	5	6	7	8	9	10	11	12	13	14
[a:]	M	0.83	0.94	0.85	0.94	0.83	0.86	0.90	0.87	0.87	0.83	0.89	0.95	0.90	0.89
	SD	0.02	0.02	0.04	0.02	0.06	0.02	0.03	0.1	0.04	0.03	0.03	0.01	0.03	0.01
								0.88 ± 0.04							
[i:]	M	0.88	0.90	0.91	0.96	0.86	0.84	0.82	0.92	0.83	0.90	0.91	0.92	0.88	0.90
	SD	0.05	0.02	0.02	0.02	0.05	0.03	0.07	0.02	0.03	0.03	0.06	0.03	0.03	0.05
								0.89 ± 0.04							
[ə:]	M	0.90	0.92	0.82	0.95	0.90	0.87	0.78	0.87	0.90	0.89	0.79	0.95	0.81	0.88
	SD	0.02	0.01	0.05	0.01	0.01	0.01	0.02	0.03	0.04	0.08	0.03	0.01	0.08	0.04
								0.87 ± 0.05							

sEMG signals across the different phonating tasks. Additionally, a few overflows were observed to be dispersed away from the central areas in phonating [a:] and [ə:] vowels.

To further confirm if the correlations between the speech signals and the sEMG recordings are consistent across all of the individuals, the correlation coefficients between the nMF and nRMS values for all the 14 subjects were obtained when [a:], [i:], and [ə:] vowels were phonated from a low to a high pitch (**Table 1**). The results presented in **Table 1** are expressed in terms of the mean and standard deviation values of the correlation coefficients across the three repeated tests for all 14 subjects. It could be observed from **Table 1** that for the [a:] vowel, an average correlation coefficient of 0.88 ± 0.04 ($p < 0.05$), was achieved across subjects while a maximum correlations value of 0.95 and was recorded by subject 12, indicating a significant correlation between the nRMS and the nMF when phonating vowel [aː]. A mean correlation coefficient of ∼0.89 ± 0.04 was recorded across subjects when the vowel [iː] was phonated, and of almost 0.87 ± 0.05 when phonating vowel [əː], with 0.96 as the highest value (subject 4) also indicating significant interrelations between the nMF of the speech signals and the nRMS of HD sEMG recordings when phonating [iː] vowel with a continuous increase in pitch. Similarly, when the vowel [əː] was phonated, an average correlation coefficient of 0.87 ± 0.05 was recorded across all participants with subject 5 attaining the maximum value (0.95) at $p < 0.05$, which also indicated significant correlations between the sound frequencies and muscular activities when the vowel [əː] was phonated.

DISCUSSION

The principal objective of this study was to investigate the interrelations between the pitches and the corresponding muscular activities when phonating different vowels with increasing pitches, which would be helpful for evaluating the pitch-related phonating functions. This objective has been reached by dynamically visualizing the myoelectrical activities of the facial/neck muscles and quantitatively investigating the correlations between sound frequencies and myoelectric characteristics during different phonating tasks.

In this study, the speech and the HD sEMG signals were synchronously acquired from 14 subjects when phonating [aː], [iː], and [əː] vowels with increasing pitches. The synchronous analysis of the speech and HD sEMG signals would clarify the interrelations between the pitches and the myoelectric activities of facial and neck muscles associated with pitch-related phonation. The sEMG signals display the potential difference between two separate electrodes above the muscles on the skin, which can be used to assess muscular function by recording electric potential activities generated by muscle cells that are from the inherent and extrinsic muscles. Moreover, phonating is a complex process controlled by multiple articulatory muscles, which contain the inherent and extrinsic muscles spanning a relatively large area of the face and neck. There were ∼30 facial and neck muscles that are involved in a phonating movement. Therefore, a total of 120 channels of closely spaced electrodes were used to construct the 2D electrode arrays for recording the HD sEMG signals that were processed in the space dimension to assess global facial and neck muscle activities corresponding to pitch-related phonation. The use of the HD sEMG signals, which provided spatiotemporal information on muscular activities, could address the limitations that finite myoelectric information obtained from several muscles, which might not afford enough functional characteristics for dynamic assessment of the articulatory muscles associated with an entire phonation process. As it is known, the nMF of the signals is defined as the frequency value at which the signal power spectrum was divided into two sections of equal energy content, and it can afford an acceptably good representation of the sound frequency shift. Meanwhile, the nRMS values of the sEMG signals represent the average power of the myoelectric activities (Phinyomark et al., 2012). Therefore, the two parameters, nMF of the speech signals and the nRMS of the sEMG recordings, were extracted to quantitatively evaluate the correlations between sound frequencies and myoelectric features associated with pitch-related phonation.

During the phonating process, the characteristics of the speech signals demonstrated that the sound frequencies were increasing with the continuous rising of the pitches while the sound volume was kept on a stable scale (**Figures 2**, **3**). These results indicated that the sound frequencies were tracked with the pitches, which was corresponding to the previous studies where the pitch was used to order sounds on a frequency-related scale (Lolli et al., 2015). In the current study, the results of the sound volume indicated that the effects on the muscle activities came from the variation of the sound frequency rather than the loudness.

Then, the temporal waveforms of HD sEMG signals across all the 120 channels on the facial and neck regions displayed that the myoelectric activities of the facial and neck muscles were orderly altered with the variation of the sound pitches (**Figure 4**). The sEMG waveforms from one electrode were different from each other even if it was from the two adjacent electrodes, which reveals that there was no cross-talk of the HD sEMG signals. The results suggested that the facial and neck muscles could produce synergetic muscle contractions to aid the movement of the passive and active articulators for constructing sound in high intelligibility with a wide scale of the pitch.

In this study, the typical normal phonating process (pitch rising) was divided into a sequence of duration, and the RMS values of the HD sEMG signals in each duration were calculated and visualized as pixels placed at a certain location of the electrode array on the facial and neck muscles. Thus, a series of 2D temporal sEMG energy maps of the facial and neck muscles were constructed for displaying the dynamic activities of the facial and neck muscles when phonating different vowels [(aː) in **Figure 5**, (iː) in **Figure 7**, and (əː) in **Figure 8**]. Also, the spatial and temporal properties of the myoelectric activities were dynamically presented in the HD sEMG energy maps to visualize the muscular contraction patterns corresponding to the pitch-related phonation activities.

According to the HD sEMG energy maps, it was found that the facial and neck muscles were both activated during phonating tasks, and the strength of the facial and neck muscles gradually increased while the high-density areas of these muscles progressively enlarged when the pitches were continuously rising. These spatiotemporal properties of the facial and neck muscles corresponded to the physiological process of phonation as a transmission path of the airstream used for generating certain sounds, during which the airstream was expelled from the lungs through the throat (controlled by the neck muscles) into the mouth (controlled by the facial muscles) (Miller et al., 2009). Meanwhile, the sound pitches were determined by the oscillatory rates of the vocal cords with orderly opening and closing, which were controlled by the facial and neck muscles that were synergistically operated with alternate contracting and stretching in precision (Echternach et al., 2017). These coordinated sequences of the facial and neck muscles would produce myoelectric activities related to phonation activities. Moreover, when the sound pitches continuously increased from low to high, the vocal cords needed to vibrate faster and faster to reach the high pitches. During this pitch-increasing process, it could require more articulatory muscles and higher muscle strength to control the airstream transmitting faster for generating the faster vibration of the vocal cords, thus resulting in the higher intensities and larger areas of muscle activities on the face and neck. Additionally, the results of the energy maps also displayed that the intensities of the neck muscles were higher than that of the facial muscles, which reveals that the neck muscles would be the main contributors during phonating movements. The signals from the middle region of the neck demonstrated larger amplitude when compared with other regions, as could be observed from the columns 6–11 in **Figure 4**. When observing the energy distribution in **Figure 5**, similar consistent results were found for the activation patterns of the articulatory muscles with larger activities in the middle region of the neck in **Figure 5A** and middle frames in **Figure 5B**. This phenomenon is also based on the physiological characteristics of the phonating process, in which the vocal cords located inside of the larynx are the major organ to produce a natural voice, whose activities are controlled by the middle neck muscles.

Furthermore, the energy concentration was symmetrically distributed on the left and right sides of the facial and larynx regions, which indicated that the energy of muscles associated with phonation activities had a nearly equivalent intensity and durability on the right and left sides of the facial and neck muscles. Notably, there was a little difference between the left and right facial muscles, such as the intensities of the left facial muscles were a bit higher than that of the right one during phonation. This phenomenon would be based on personal habits in speaking and swallowing. For instance, if someone eats bolus by often using the left side of the teeth, the left facial muscles would be more exercised than the right one, which would lead to stronger muscles on the left face. Thus, when the subject phonated different vowels, the activities of the left facial muscles were higher than that of the right one. This result suggested that the HD sEMG energy maps might be a useful tool for guiding the subjects to rectify their undesirable habits in speaking and swallowing activities.

The HD energy maps obtained from different phonating tasks showed the effects of the facial and neck muscles across different vowels [(a:), (i:), and (ə:)]. Based on the HD energy maps, the high intensity of the maps on the neck region appeared before that of the facial areas, which indicated that the muscular activities from the neck muscles were activated before the facial muscles at the onset of the phonating [a:] and [ə:] vowels (**Figures 5**, **8**). In contrast, the activities of the neck muscles presented a little more delay than the facial muscles when phonating the [i:] vowel (**Figure 7**). These results indicated that the styles of the pronunciation would produce different voices, which required sequential contractions of facial and neck muscles in a certain order. Moreover, the energy maps on the facial muscles showed that the high-intensity areas appeared on different muscles when phonating different vowels. For instance, when phonating back vowel [a:], the subject should first relax and open the mouth, while keeping the lips in a round shape; this manner of pronunciation would lead to the strong strength of the masseter muscles. Hence, the high intensities of the maps were concentrated on the edge of the face both on the left and right sides (**Figure 5B**). On the contrary, when phonating the front vowel [i:], the subject needed to extend the lips to the sides in a flat shape, which could cause the activation of the orbicularis oris muscles, resulting in the high-intensity activities around the mouth (**Figure 7**). Finally, when phonating the central vowel [i:], the mouth of the subject just needed to be opened slightly, and the lips should keep in relaxation. Thus, the strength of facial muscles was very weak, and there were low intensities presented in the energy maps of the facial regions (**Figure 8**). Moreover, no significant effects of the phonating task order on the HD sEMG results were found in our pilot experiments of this study.

Consequently, the dynamic properties of the HD sEMG energy maps were well corresponding to the physiological and biomechanical principles of phonation, which would make the HD sEMG method a simple and non-invasive tool for accurately visualizing the phonating process and assessing the temporal and spatial properties of facial and neck muscles associated with phonation functions. Moreover, it could also help researchers and laryngologists to better understand the dynamic muscular activities of the facial and neck muscles during phonation. The HD sEMG technique might pave the way for developing a clinically relevant approach to screen, diagnose the phonating disorder caused by muscle problems, and even provide a potential method to locate the abnormal muscles that lead to the phonating disorder. Furthermore, the findings from the proposed method are not only applicable to the assessment of phonating function but may also spur positive advancement in other rehabilitation areas, including swallowing function evaluation (Zhu et al., 2017b) and back pain rehabilitation (Hu et al., 2014), among others, that equally adopts sEMG maps.

Considering the fact that phonation is a complex process controlled by multiple articulatory muscles, the use of both speech signals and sEMG recordings are necessary indicators for the analysis of the pitch-related phonation activities (Zhu et al., 2017b; Murtola et al., 2019; Zhang et al., 2020). However, it should be noted that the correlations between the sound frequencies (pitches) and the muscular activities of facial and neck muscles during a pitch-related phonation remains unclear. In this investigation, the nMF of the speech signals and the nRMS of the HD sEMG recordings were measured and compared during a pitch increase across three different phonating tasks [(a:), (i:), and (ə:) vowels]. The results presented that the nRMS values (from sEMG signals) increased at a similar rate with the nMF values (from speech signals) when constantly increasing pitches (**Figures 6**, **9**). These results demonstrated that a strong interrelation existed between the sound frequencies and the muscular activities associated with pitch-related phonation. Meanwhile, the results of the correlation coefficients showed that there were significant correlations between the nMF of speech signals and the nRMS of sEMG recordings across different vowel phonation tasks [back (a:), front (i:), and central (ə:)] (**Table 1**).

The findings revealed that the muscular activity patterns of the facial and neck muscles related to phonation might be served as a reference for evaluating the pitch-related phonation functions.

The purpose of the present study focused on the normal phonating functions, so the patients with phonating disorders will be recruited for further study. The HD energy maps and the interrelations between the speech signals and the sEMG recordings obtained in this investigation might be used as the reference for further evaluating and diagnosing the phonating functions or dysphonia in clinical application.

CONCLUSIONS

In this study, the speech and the HD sEMG signals were simultaneously measured when the subjects were phonating vowels with an increasing pitch. The sequential energy maps constructed from multichannel sEMG recordings showed that muscle contraction strength increased monotonously as the pitch increased, and left-right symmetrical distribution was observed for the face and neck regions when phonating different vowels with increasing pitch. The nRMS parameter of the HD sEMG signals increased in a similar pattern to the nMF parameter of the speech signals when the pitch rose, and there were significant correlations between the two parameters across different vowel tasks. The present study suggested that the muscle contraction patterns might be used as a reference for the evaluation of pitch-related phonation functions that is important for dysphonia diagnoses. This study might also spur positive advancement in the evaluation of facial paralysis and diagnoses of other neuromuscular-inclined diseases.

DATA AVAILABILITY STATEMENT

The data used to support the findings of this study are available on reasonable request from the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board of Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences (SIAT-IRB-170815-H0178). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SC and GL: conceptualization. MZ and XW: data curation and writing—original draft. MZ, SC, and GL: funding acquisition. MZ and HD: methodology. SC: supervision. XW, YH, and HZ: validation. MZ and ZL: visualization. MZ and MW: writing—review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the National Natural Science Foundation of China (#81927804, #62101538, and #61901464), the Shenzhen Governmental Basic Research Grant (#JCYJ20180507182241622), the Science and Technology Planning Project of Shenzhen (#JSGG20210713091808027 and #JSGG20211029095801002), and the SIAT Innovation Program for Excellent Young Researchers (E1G027).

REFERENCES

Alku, P. (2011). Glottal inverse filtering analysis of human voice production — a review of estimation and parameterization methods of the glottal excitation and their applications. *Sadhana* 36, 623–650. doi: 10.1007/s12046-011-0041-5

Balata, P. M. M., da Silva, H. J., de Araújo Pernambuco, L., de Oliveira, J. H. P., and de Moraes, S. R. A. (2015). Normalization patterns of the surface electromyographic signal in the phonation evaluation. *J. Voice* 29, e121.1–8. doi: 10.1016/j.jvoice.2014.03.010

Bishop, J., and Keating, P. (2012). Perception of pitch location within a speaker's range: fundamental frequency, voice quality and speaker sex. *J. Acoust. Soc. Am.* 132, 1100–1112. doi: 10.1121/1.4714351

Bracken, D. J., Ornelas, G., Coleman, T. P., and Weissbrod, P. A. (2019). High-density surface electromyography: a visualization method of laryngeal muscle activity. *Laryngoscope* 129, 2347–2353. doi: 10.1002/lary.27784

Brajot, F.-X., and Lawrence, D. (2018). Delay-induced low-frequency modulation of the voice during sustained phonation. *J. Acoust. Soc. Am.* 144, 282–291. doi: 10.1121/1.5046092

Chen, M., Zhang, X., and Zhou, P. (2018). A novel validation approach for high-density surface EMG decomposition in motor neuron disease. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 1161–1168. doi: 10.1109/TNSRE.2018.2836859

Chhetri, D. K., and Neubauer, J. (2015). Differential roles for the thyroarytenoid and lateral cricoarytenoid muscles in phonation. *Laryngoscope* 125, 2772–2777. doi: 10.1002/lary.25480

Craig, J., Tomlinson, C., Stevens, K., Kotagal, K., Fornadley, J., Jacobson, B., et al. (2015). Combining voice therapy and physical therapy: a novel approach to treating muscle tension dysphonia. *J. Commun. Disord.* 58, 169–178. doi: 10.1016/j.jcomdis.2015.05.001

Cutiva, L. C. C., Vogel, I., and Burdorf, A. (2013). Voice disorders in teachers and their associations with work-related factors: a systematic review. *J. Commun. Disord.* 46, 143–155. doi: 10.1016/j.jcomdis.2013.01.001

Dennis, J., Tran, H. D., and Li, H. (2010). Spectrogram image feature for sound event classification in mismatched conditions. *IEEE Signal Process. Lett.* 18, 130–133. doi: 10.1109/LSP.2010.2100380

Dewan, K., Vahabzadeh-Hagh, A., Soofer, D., and Chhetri, D. K. (2017). Neuromuscular compensation mechanisms in vocal fold paralysis and paresis. *Laryngoscope* 127, 1633–1638. doi: 10.1002/lary.26409

Echternach, M., Burk, F., Köberlein, M., Herbst, C. T., Döllinger, M., Burdumy, M., et al. (2017). Oscillatory characteristics of the vocal folds across the tenor passaggio. *J. Voice* 31, 381.e5–e381.e14. doi: 10.1016/j.jvoice.2016.06.015

Gerratt, B. R., Kreiman, J., and Garellek, M. (2016). Comparing measures of voice quality from sustained phonation and continuous speech. *J. Speech Lang. Hear. Res.* 59, 994–1001. doi: 10.1044/2016_JSLHR-S-15-0307

Glaser, V., and Holobar, A. (2018). Motor unit identification from high-density surface electromyograms in repeated dynamic muscle contractions. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 66–75. doi: 10.1109/TNSRE.2018.2885283

Horáček, J., Bula, V., Radolf, V., Vampola, T., and Duškov,á, M. (2016). Development of self-oscillating human vocal folds prosthesis. *Proc. Eng.* 144, 867–874. doi: 10.1016/j.proeng.2016.05.103

Hu, Y., Kwok, J. W., Tse, J. Y.-H., and Luk, K. D.-K. (2014). Time-varying surface electromyography topography as a prognostic tool for chronic low back pain rehabilitation. *Spine J.* 14, 1049–1056. doi: 10.1016/j.spinee.2013.11.060

Jani, M. P., and Gore, G. B. (2014). Occurrence of communication and swallowing problems in neurological disorders: analysis of forty patients. *NeuroRehabilitation* 35, 719–727. doi: 10.3233/NRE-141165

Johns, G., Morin, E., and Hashtrudi-Zaad, K. (2016). Force modelling of upper limb biomechanics using ensemble fast orthogonal search on high-density electromyography. *IEEE Trans. Neural Syst. Rehabil. Eng.* 24, 1041–1050. doi: 10.1109/TNSRE.2016.2515087

Kaneko, M., Hitomi, T., Takekawa, T., Tsuji, T., Kishimoto, Y., and Hirano, S. (2018). Effects of voice therapy on laryngeal motor units during phonation in chronic superior laryngeal nerve paresis dysphonia. *J. Voice* 32, 729–733. doi: 10.1016/j.jvoice.2017.08.026

Khoddami, S. M., Talebian, S., Izadi, F., and Ansari, N. N. (2017). Validity and reliability of surface electromyography in the assessment of primary muscle tension dysphonia. *J. Voice* 31, 386.e9–386.e17. doi: 10.1016/j.jvoice.2016.09.010

Kim, J., Wigram, T., and Gold, C. (2009). Emotional, motivational and interpersonal responsiveness of children with autism in improvisational music therapy. *Autism* 13, 389–409. doi: 10.1177/1362361309105660

Knuijt, S., Kalf, J. G., de Swart, B. J., Drost, G., Hendricks, H. T., Geurts, A. C., et al. (2014). Dysarthria and dysphagia are highly prevalent among various types of neuromuscular diseases. *Disabil. Rehabil.* 36, 1285–1289. doi: 10.3109/09638288.2013.845255

Kryshtopava, M., Van Lierde, K., Meerschman, I., D'haeseleer, E., De Moor, M., Vandemaele, P., et al. (2017). Functional magnetic resonance imaging study of brain activity associated with pitch adaptation during phonation in healthy women without voice disorders. *J. Voice* 31, 118.e21–e28. doi: 10.1016/j.jvoice.2016.02.022

Lolli, S., Lewenstein, A. D., Basurto, J., Winnik, S., and Loui, P. (2015). Sound frequency affects speech emotion perception: results from congenital amusia. *Front. Psychol.* 6, 1340. doi: 10.3389/fpsyg.2015.01340

Macdonald, I., Rubin, J. S., Blake, E., Hirani, S., and Epstein, R. (2012). An investigation of abdominal muscle recruitment for sustained phonation in 25 healthy singers. *J. Voice* 26, 815.e19–815.e16. doi: 10.1016/j.jvoice.2012.04.006

Martins, R. H. G., Pereira, E. R. B. N., Hidalgo, C. B., and Tavares, E. L. M. (2014). Voice disorders in teachers. A review. *J. Voice* 28, 716–724. doi: 10.1016/j.jvoice.2014.02.008

Miller, A. L., Brugman, J., Sands, B., Namaseb, L., Exter, M., and Collins, C. (2009). Differences in airstream and posterior place of articulation among N| uu clicks. *J. Int. Phon. Assoc.* 39, 129–161. doi: 10.1017/S0025100309003867

Murtola, T., Malinen, J., Geneid, A., and Alku, P. (2019). Analysis of phonation onsets in vowel production, using information from glottal area and flow estimate. *Speech Commun.* 109, 55–65. doi: 10.1016/j.specom.2019.03.007

Naik, G. R., Al-Timemy, A. H., and Nguyen, H. T. (2015). Transradial amputee gesture classification using an optimal number of sEMG sensors: an approach using ICA clustering. *IEEE Trans. Neural Syst. Rehabil. Eng.* 24, 837–846. doi: 10.1109/TNSRE.2015.2478138

Pettersen, V., Bjørkøy, K., Torp, H., and Westgaard, R. H. (2005). Neck and shoulder muscle activity and thorax movement in singing and speaking tasks with variation in vocal loudness and pitch. *J. Voice* 19, 623–634. doi: 10.1016/j.jvoice.2004.08.007

Phinyomark, A., Phukpattaranont, P., and Limsakul, C. (2012). Feature reduction and selection for EMG signal classification. *Expert Syst. Appl.* 39, 7420–7431. doi: 10.1016/j.eswa.2012.01.102

Rinaldi, L., Lega, C., Cattaneo, Z., Girelli, L., and Bernardi, N. F. (2016). Grasping the sound: auditory pitch influences size processing in motor planning. *J. Exp. Psychol. Hum. Percept. Perform.* 42, 11. doi: 10.1037/xhp0000120

Rosen, C. A., and Murry, T. (2000). Voice handicap index in singers. *J. Voice* 14, 370–377. doi: 10.1016/S0892-1997(00)80082-X

Sommerville, R. B., Vincenti, M. G., Winborn, K., Casey, A., Stitziel, N. O., Connolly, A. M., et al. (2017). Diagnosis and management of adult hereditary cardio-neuromuscular disorders: a model for the multidisciplinary care of complex genetic disorders. *Trends Cardiovasc. Med.* 27, 51–58. doi: 10.1016/j.tcm.2016.06.005

Strazzulla, I., Nowak, M., Controzzi, M., Cipriani, C., and Castellini, C. (2016). Online bimanual manipulation using surface electromyography and incremental learning. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 227–234. doi: 10.1109/TNSRE.2016.2554884

Tang, X., Zhang, X., Gao, X., Chen, X., and Zhou, P. (2018). A novel interpretation of sample entropy in surface electromyographic examination of complex neuromuscular alternations in subacute and chronic stroke. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 1878–1888. doi: 10.1109/TNSRE.2018.2864317

Van Houtte, E., Claeys, S., D'haeseleer, E., Wuyts, F., and Van Lierde, K. (2013). An examination of surface EMG for the assessment of muscle tension dysphonia. *J. Voice* 27, 177–186. doi: 10.1016/j.jvoice.2011.06.006

Xu, X., Yang, P., Zhuang, P., Yanchao, J., Yanli, M., Schrof, C., et al. (2018). Study on normal laryngeal electromyography of thyroarytenoid muscle, cricothyroid muscle, and posterior cricoarytenoid muscle. *Ann. Otol. Rhinol. Laryngol.* 127, 806–811. doi: 10.1177/0003489418796525

Zhang, M., Wang, Y., Wei, Z., Yang, M., Luo, Z., and Li, G. (2020). Inductive conformal prediction for silent speech recognition. *J. Neural Eng.* 17:066019. doi: 10.1088/1741-2552/ab7ba0

Zhu, M., Liang, F., Samuel, O. W., Chen, S., Yang, W., Lu, L., et al. (2017a). "A pilot study on the evaluation of normal phonating function based on high-density sEMG topographic maps," in *the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jeju, 1030–1033. doi: 10.1109/EMBC.2017.8037002

Zhu, M., Lu, L., Yang, Z., Wang, X., Liu, Z., Wei, W., et al. (2018a). "Contraction patterns of neck muscles during phonating by high-density surface electromyography," in *the 2018 IEEE International Conference on Cyborg and Bionic Systems (CBS)*, Shenzhen, 572–575. doi: 10.1109/CBS.2018.8612181

Zhu, M., Samuel, O. W., Yang, Z., Lin, W., Huang, Z., Fang, P., et al. (2018b). "Using muscle synergy to evaluate the neck muscular activities during normal swallowing," in *the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Honolulu, HI, 2454–2457. doi: 10.1109/EMBC.2018.8512760

Zhu, M., Yu, B., Yang, W., Jiang, Y., Lu, L., Huang, Z., et al. (2017b). Evaluation of normal swallowing functions by using dynamic high-density surface electromyography maps. *Biomed. Eng.* 16, 133. doi: 10.1186/s12938-017-0424-x

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

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 26 July 2022
DOI 10.3389/fnins.2022.971039

OPEN ACCESS

EDITED BY
Xiaomao Fan,
South China Normal University, China

REVIEWED BY
Wenjun Tan,
Northeastern University, China
Mengqi Shen,
Virginia Tech, United States

*CORRESPONDENCE
Longlong Cheng
chenglonglong@cecdat.com
Chao Chen
cccovb@hotmail.com

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

RECEIVED 16 June 2022
ACCEPTED 07 July 2022
PUBLISHED 26 July 2022

CITATION
Li P, Su J, Belkacem AN, Cheng L and Chen C (2022) Multi-person feature fusion transfer learning-based convolutional neural network for SSVEP-based collaborative BCI.
Front. Neurosci. 16:971039.
doi: 10.3389/fnins.2022.971039

Multi-person feature fusion transfer learning-based convolutional neural network for SSVEP-based collaborative BCI

Penghai Li[1], Jianxian Su[1], Abdelkader Nasreddine Belkacem[2], Longlong Cheng[3]* and Chao Chen[4]*

[1]School of Integrated Circuit Science and Engineering, Tianjin University of Technology, Tianjin, China, [2]Department of Computer and Network Engineering, College of Information Technology, UAE University, Al Ain, United Arab Emirates, [3]China Electronics Cloud Brain Technology Co., Ltd., Tianjin, China, [4]Key Laboratory of Complex System Control Theory and Application, Tianjin University of Technology, Tianjin, China

Objective: The conventional single-person brain–computer interface (BCI) systems have some intrinsic deficiencies such as low signal-to-noise ratio, distinct individual differences, and volatile experimental effect. To solve these problems, a centralized steady-state visually evoked potential collaborative BCI system (SSVEP-cBCI), which characterizes multi-person electroencephalography (EEG) feature fusion was constructed in this paper. Furthermore, three different feature fusion methods compatible with this new system were developed and applied to EEG classification, and a comparative analysis of their classification accuracy was performed with transfer learning-based convolutional neural network (TL-CNN) approach.

Approach: An EEG-based SSVEP-cBCI system was set up to merge different individuals' EEG features stimulated by the instructions for the same task, and three feature fusion methods were adopted, namely parallel connection, serial connection, and multi-person averaging. The fused features were then input into CNN for classification. Additionally, transfer learning (TL) was applied first to a Tsinghua University (THU) benchmark dataset, and then to a collected dataset, so as to meet the CNN training requirement with a much smaller size of collected dataset and increase the classification accuracy. Ten subjects were recruited for data collection, and both datasets were used to gauge the three fusion algorithms' performance.

Main results: The results predicted by TL-CNN approach in single-person mode and in multi-person mode with the three feature fusion methods were compared. The experimental results show that each multi-person mode is superior to single-person mode. Within the 3 s time window, the classification accuracy of the single-person

CNN is only 90.6%, while the same measure of the two-person parallel connection fusion method can reach 96.6%, achieving better classification effect.

Significance: The results show that the three multi-person feature fusion methods and the deep learning classification algorithm based on TL-CNN can effectively improve the SSVEP-cBCI classification performance. The feature fusion method of multi -person parallel feature connection achieves better classification results. Different feature fusion methods can be selected in different application scenarios to further optimize cBCI.

KEYWORDS

steady-state visually evoked potential, collaborative BCI, feature fusion, convolutional neural network, transfer learning

Introduction

Brain–computer interface (BCI) is a human–computer interaction technology that allows people to directly communicate with a computer or control peripheral device without their surrounding muscles (Vaid et al., 2015). This technology is useful for patients with movement disorders and partial brain injuries, as it helps them realize simple operation and communication (Wolpaw et al., 2000). At present, electroencephalography (EEG)-BCI systems mainly include event-related potentials evoked by endogenous events based on cognitive function (Li et al., 2019), visually evoked potentials (VEP) based on visual stimulation (Mary Judith and Baghavathi Priya, 2021), and event-related area synchronization and event-based active motor imagery in the phenomenon of correlation synchronization (Munzert et al., 2009). Steady-state visually evoked potential (SSVEP) is one of the most popular EEG patterns in the field of BCI. Owing to its advantages such as high information transmission rate (ITR), low requirement on user training, and easy evocation, SSVEP is widely applied to various fields such as medical care, industries, communication, smart home, gaming, robotics, and vehicle control (Zhao et al., 2016; Angrisani et al., 2018; Dehzangi and Farooq, 2018; Farmaki et al., 2019; Nayak et al., 2019; Chai et al., 2020; Shao et al., 2020).

Single-person BCI system's performance is subject to individual differences between users and their physical or mental conditions, and this weakness becomes more prominent as BCI system develops further (Song et al., 2022). In contrast, multi-person-coordinated BCI can better serve the future socialized human–computer interaction and will most certainly dominate this field both in terms of research and application. Studies have shown that increasing the number of users can substantially improve BCI performance (Valeriani et al., 2016). In human behavior research, teams' performance is always better than that of individuals. The distinction in performance between teams and individuals is even greater when humans acquire diversified skills, judgments, and experiences under time constraints (Katzenbach and Smith, 2015). As single-person EEG signals have significant individual differences, by collecting multi-person EEG signals and fusing these signals in a reasonable way, signals with more distinctive features can be obtained, and the BCI performance can be improved. EEG signals from multiple subjects can significantly improve ITR in the system compared to single EEG signals (Bianchi et al., 2019). Subjects who need to stare at the stimulation area for a long time are prone to fatigue due to visual stimulation in SSVEP-BCI, which affects the quality of EEG signal acquisition, and this is particularly evident for some subjects (Peng et al., 2019). SSVEP-cBCI can make up for this deficiency by increasing the user dimension and improve the information transmission rate. Acknowledging this viewpoint, this paper explores three feature fusion methods, which include (1) parallel connecting features, (2) serial concatenating features, and (3) feature averaging. These approaches will be explained in detail in section "Methods." The three feature fusion methods aim to improve the signal-to-noise ratio by merging multi-person EEG information to get refined new features to enhance the BCI performance.

As a branch of machine learning, deep learning has achieved great success in solving problems in computer vision and natural language processing. It is different from traditional machine learning as it does not entail manual feature extraction (LeCun et al., 2015). Using gradient descent learning to optimize convolutional neural network (CNN) parameters successfully solved the problem of handwritten digit classification (LeCun et al., 1998). However, owing to the complexity of EEG signals, the application of deep learning neural networks in EEG signal detection is still in the exploratory stage. Cecotti and Graser (2010) developed a four-layer CNN for P300 detection. At present, the SSVEP EEG signal classification method converts the original EEG signal through FFT and then inputs it into CNN for classification (Cecotti, 2011; Zhang et al., 2019; Ravi et al., 2020). As a superb CNN model designed for EEG, EEGNet exhibits good classification performance, but other models perform better in some moments. In this study, some details of the basic EEGNet were adjusted, and the network structure was modified to adapt to the newly created fusion features. The transfer learning (TL) training strategy using a THU benchmark dataset as the source task training set was adopted to initially train the parameters of the convolutional layer and build the basic feature extractor. Using the data collected by the laboratory as the target task training set and test set, the CNN parameters were further optimized to construct SSVEP-cBCI. In this paper, the classification model is trained with the TL-CNN method, which reduces the required amount of training collected data and improves the classification accuracy. And the feature fusion approach further improves BCI performance in classification accuracy, ITR and stability.

Section "Methods" elaborates on the personnel, equipment, and experimental paradigms associated with the experiments, the three multi-person features fusion methods, the specific structure of the modified CNN in this study, and its difference from EEGNet. Then the following part introduces the specific training method of TL. In section "Results," the classification accuracy and ITR difference of the three feature fusion methods and those predicted by a single-person CNN are compared. Finally, some significant conclusions are drawn, and the specific usage of the three feature fusion methods in this experiment is analyzed.

Methods

Experimental setup

The structure of cBCI system

The cBCI system mainly has two structural forms: distributed and centralized (Wang and Jung, 2011). In both systems, experiments are simultaneously conducted on more than one subject. In the distributed cBCI, subjects' EEG information is collected individually for subsequent data preprocessing, feature extraction, and pattern recognition through the corresponding BCI subsystem. The results corresponding to each subject are then transmitted to the integrated classifier, and the final decision is produced through decision-making layer's voting mechanism, while in the centralized cBCI, as shown in **Figure 1**, subjects' EEG information is collected individually for sequential data preprocessing and feature extraction. The EEG data features of all subjects are fused for pattern recognition to make the final decision for the group. The model adopted in this study is a centralized cBCI system, which does not rely on the voting mechanism of the distributed system, and classification is carried out with a CNN based on TL (TL-CNN).

Experimental paradigm

In this experiment, the EEG data were collected and transferred from the EEG amplifier to the software Curry8 (Neuroscan). Three electrodes were placed on O1, Oz, and O2 according to the International 10–20 system. Using the double mastoid as reference and ground electrodes, the impedance of all electrodes was reduced to below 5 kΩ. The sampling frequency is 256 Hz, and a band-pass filter between 5 and 40 Hz is used in the data processing to filter out low-frequency noise and 50 Hz power frequency noise.

Ten healthy subjects (8 males, 2 females, 21–27 years old) participated in the experiments. All participants had normal or corrected vision. Four of them had participated in SSVEP experiments previously. All participants read and signed the informed consent forms. Subjects sat on a comfortable chair 60 cm in front of a standard 24-inch monitor (60 Hz refresh rate, 1,920 × 1,080 screen resolution). The SSVEP stimulation interface is shown in **Figure 2**, and the four stimulation squares are all 50 × 50 pixels. The refresh frequency of the display equals integer multiples of the stimulation frequency of the four color blocks, which can ensure stable stimulation frequency and avoid frequency deviation. The stimulation frequencies of the four color blocks are 8.6, 10, 12, and 15 Hz, respectively. It was evidenced that stimulation frequencies of 10 and 12 Hz can stably induce high-amplitude SSVEP signals (Chen et al., 2015), and the stimulation duration was set to be 4 s. To avoid interference caused by simultaneous flickering of the four color blocks, the phases of the four color blocks are set as 1.35π, 0.35π, 0.9π, and 0.35π, respectively. Prolonged staring at the flickering stimulus color blocks made the subjects feel tired and distracted them, resulting in a frequency deviation of the SSVEP signal. To improve the concentration of the subjects and the quality of SSVEP EEG signals, random labels were used to remind the subjects to look at the corresponding stimulus squares.

0.02 s after the five-pointed star appeared, the four color blocks started to flash. After the flashing, a rest time of 2 s was given to the subject to adjust the viewing angle. During the experiment, the subjects were asked to focus on the corresponding color block and blink as few times as possible. Each color block flashed twice in total, and there was a 1-min rest between two consecutive experiments.

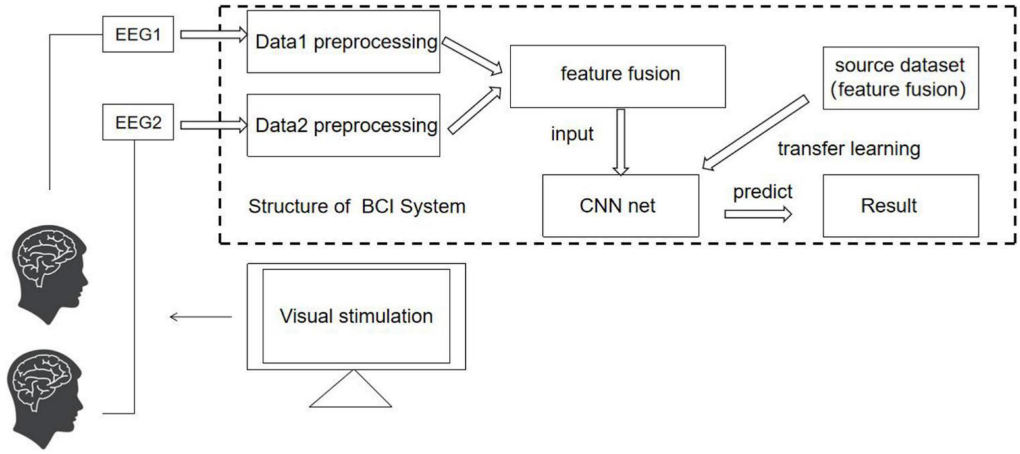


FIGURE 1
Centralized cBCI structure designed in this study.

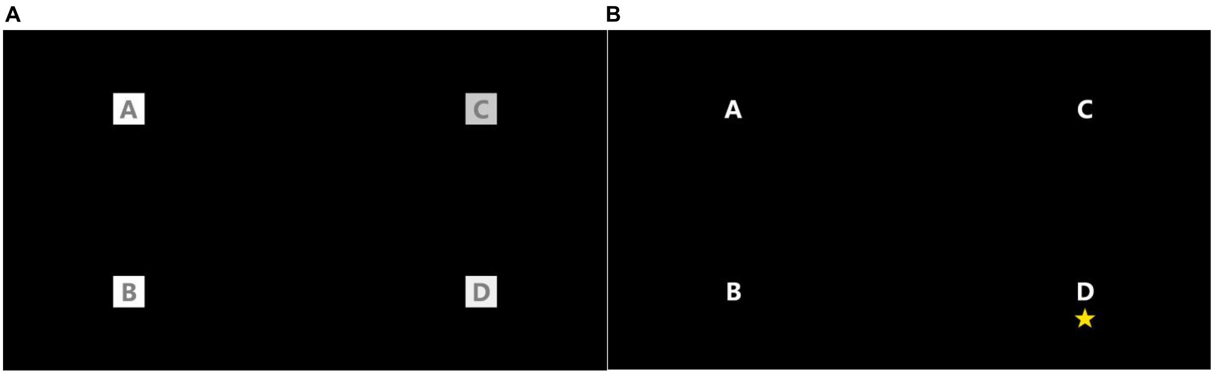


FIGURE 2
SSVEP stimulation interface and label reminder method. **(A)** Stimulus interface. **(B)** Random tag prompt.

Multi-Person feature extraction and fusion

Multi-Person feature extraction

The EEG data filtered and processed by the fourth-order Butterworth filter is converted from the time domain to the frequency domain by FFT transformation (Chen et al., 2015). Low-frequency (8.6, 10, 12, and 15 Hz) stimulation area was used in these experiments. The features of the frequency band from 6 to 32 Hz were selected from the FFT-transformed data to further filter out noise and improve feature quality.

The characteristics of the SSVEP signal are as follows:

$$\begin{cases} Featue_{O1} = |FFT(X_{O1})| \\ Featue_{OZ} = |FFT(X_{Oz})| \\ Featue_{O2} = |FFT(X_{O2})| \end{cases} \tag{1}$$

The input of the convolutional neural network is:

$$Input = \begin{vmatrix} min\_max(Featue_{O1}) \\ min\_max(Featue_{Oz}) \\ min\_max(Featue_{O2}) \end{vmatrix} \tag{2}$$

The min–max normalization (discrete normalization) is conducted on the data of each channel (Ali et al., 2014) to avoid adverse effects on the classification accuracy owing to huge differences between values, ensure good performance of different data within the same neural network, and improve the robustness of the algorithm.

Feature fusion

This paper proposes three methods to fuse multi-person EEG features. As shown in **Figure 3**, in parallel connection, three-channel (O1, Oz, O2) data are extracted

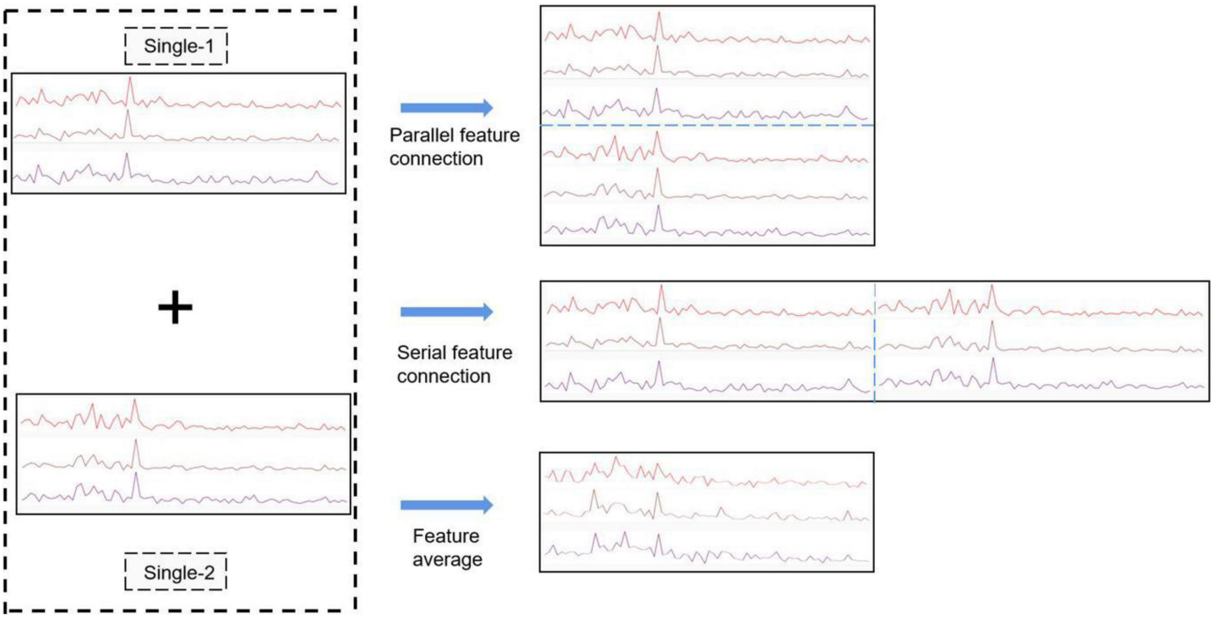


FIGURE 3
Three fusion methods: (1) feature parallel connection (2) feature serial connection (3) feature averaging. Single-1 and Single-2 represent two single-person features for feature fusion.

to obtain the effective feature after data preprocessing and FFT transformation. The features of different subjects are connected in parallel, serial concatenation, or averaging. In parallel feature concatenation, connection is made mainly the spatial domain, which implies more feature lead channels. In serial feature connection, connection is made mainly in the frequency domain, which implies that there is no change in the number of channels, but the domain scope expands greatly, and thus, the effective features are enhanced from the frequency perspective and the BCI performance improves. However, serial concatenation requires more training on epoch and convolution kernel to achieve the similar classification accuracy of parallel connection. It involves more complex algorithm, so it is more difficult to set up an online system by Python. The above two feature fusion methods are suitable for subjects with a known number of participants in the experiment, but when the number of participants in the brain group is unknown, different CNNs meant for various number of subjects should be set up and trained, which entails more input in the experimental preparation. This problem can be solved by adopting the third approach, feature averaging, that is, to get new features by averaging the normalized EEG frequency features of all subjects. The CNN using this approach shares the same structure of single-person CNN, and its classification accuracy is superior to that of a single-person CNN but inferior to that of a two-person CNN.

Deep learning network construction

Improved the network structure of convolutional neural network

Lawhern et al. (2018) designed EEGNet, a compact CNN specially for EEG signals, that is based on CNN and includes two parts: spatial feature extraction and frequency or time domain feature extraction. It can efficiently extract features from frequency-domain EEG information and send them to a neural network-based classifier, eliminating the need to manually extract two-part features. This paper makes some adjustments on the EEGNet, changing the number of convolution kernels, the size of the convolution kernel, and the depth of the convolution layer. In this experiment, the collected data were used to predict the classification accuracy of the test set, and the EEGNet was modified to accommodate the data. In this study, the ordinary convolution layer was discarded, a depthwise convolution layer was added, and the pointwise convolution layer was changed into a small narrow-band convolution. The network consists of six consecutive layers, including four convolutional layers, one fully connected layer, and one softmax output layer (Jang et al., 2016). Network fitting is accelerated through batch normalization (Ba et al., 2016). Linear activation layer adopts ReLu function (Agarap, 2018).

Table 1 summarizes the modified CNN structure. The convolution kernel of the C1 convolution layer has size 3 × 1, and its function is to learn the linear combination and spatial filtering features between different channels. The method

TABLE 1 Single-person convolutional neural network structure.

Layer number	Layer	Filter	Kernel size	Feature size	Activation
1	Input data	–	–	(3.78)	–
2	Conv2D	16	(3.1)	(3.78)	ReLU
3	Conv2D	32	(3.1)	(1.78)	ReLU
4	Conv2D	32	(1.3)	(1.78)	ReLU
5	Conv2D	64	(1.3)	(1.76)	ReLU
6	Flatten	–	–	–	–
7	Dense	8	–	–	–
8	Dropout	–	Rate = 0.5	–	–
9	Dense	4	–	–	Softmax

of padding and zero-filling is adopted to prevent the loss of information caused by convolution (Dwarampudi and Reddy, 2019). The C2 layer does not use the method of padding and zero-filling and integrates multi-channel data into a single channel by convolution. The C3 convolution layer extracts features along the input frequency spectrum by convolution and acts as a band-pass filter. The C4 convolution layer also integrates frequency features without padding. Among them, the convolution operations of the C2 and C4 layers have achieved the down-sampling effect. After Flatten layer, a fully connected layer is followed by softmax for classification.

Network structure of feature fusion convolutional neural network

As can be seen in **Table 2**, compared with the single-person CNN, the difference between the two-person parallel feature concatenation CNN structure is that it increases the number of key channels, from three-channel to six-channel EEG data, which greatly increases the number of features.

Therefore, the two-person parallel feature connection CNN structure was added to unpadded convolution layers C2 and C3 in accordance with the single-person CNN structure while keeping the fully connected layer and the last two layers unchanged. The network classification results show that the classification accuracy falls by about 1% as each of the two convolutional layers is reduced. Multiple convolution operations can effectively extract complex multi-channel features and integrate them into a single spatial feature.

TABLE 2 Two-person parallel feature connection CNN structure.

Layer number	Layer	Filter	Kernel size	Feature size	Activation
1	Input data	–	–	(6.78)	–
2	Conv2D	16	(6.1)	(6.78)	ReLU
3	Conv2D	32	(3.1)	(4.78)	ReLU
4	Conv2D	64	(3.1)	(2.78)	ReLU
5	Conv2D	64	(2.1)	(1.78)	ReLU
6	Conv2D	128	(1.3)	(1.78)	ReLU
7	Conv2D	256	(1.3)	(1.76)	ReLU
8	Flatten	–	–	–	–
9	Dense	8	–	–	–
10	Dropout	–	Rate = 0.5	–	–
11	Dense	4	–	–	Softmax

TABLE 3 Two-person serial feature connection CNN structure.

Layer number	Layer	Filter	Kernel size	Feature size	Activation
1	Input data	–	–	(3.234)	–
2	Conv2D	48	(3.1)	(3.234)	ReLU
3	Conv2D	96	(3.1)	(1.234)	ReLU
4	Conv2D	96	(1.3)	(1.234)	ReLU
5	Conv2D	192	(1.3)	(1.232)	ReLU
6	Flatten	–	–	–	–
7	Dense	8	–	–	–
8	Dropout	–	Rate = 0.5	–	–
9	Dense	4	–	–	Softmax

As can be seen in **Table 3**. The CNN structure used by the two-person serial feature connection method is similar to the single-person CNN structure. With the dual serial feature connection, the number of features input to the CNN is increased. This builds more feature extractors by increasing the number of convolution kernels to get better results. If the number of convolution kernels of the two-person CNN connected by serial features is the same as that of the single-person CNN, the classification accuracy will drop by about 2%.

Transfer learning-based feature fusion strategy with different datasets

Compared with traditional machine learning algorithms, deep learning methods heavily rely on high-quality data. Obtaining sufficient high-quality datasets to train high-quality convolution kernel parameters is a critical problem to be solved in CNN setup. Transfer learning (Pan and Yang, 2009) gives an effective solution to this problem. The SSVEP EEGs collected in the THU benchmark dataset (Chen et al., 2015) exhibit good features and low error rates of subjects' operation, and thus, this dataset was used as the source dataset for initial parameters training on the model. In general, parameters in CNN are randomly initialized by training collected data directly. Compared with transfer learning, it requires a larger amount of data and training time to fit and get a satisfactory feature extractor. While using transfer learning methods, initial parameters can be constructed in a pre-training manner, and these parameters are usually derived from prior knowledge and hence can well perform the corresponding task. As a consequence, only a small amount of actual experimental data serve as the training set, and the model parameters are re-learned through fine-tuning for the model to adapt to the actual

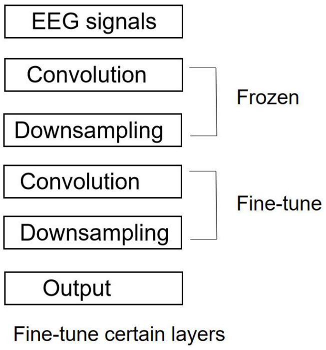


FIGURE 4
Training strategy.

experimental data. This method can improve the classification accuracy of the model and effectively reduce the required size of experimental data collected in our laboratory to train the CNN.

The comparison among various fine-tune methods suggests significant differences in their stability but insignificant difference in their classification accuracy. **Figure 4** shows that only the parameters of the deep convolutional layer and the fully connected layer are trained, while the parameters of the shallow CNN are frozen and not involved in the training. Since the feature distribution of the source task data (THU benchmark dataset) does not coincide with that in this experiment, fine-tuning on the parameters of the deep convolutional layer with a small learning rate can improve the feature extraction performance of the convolutional layer.

In the first step, training was conducted with the THU benchmark dataset as the training set, and a total of 720 samples were used, with 180 samples per stimulus. The network weights are learned in accordance with the Adam learning algorithm, which optimizes the network weights through back-propagation, and the cross-entropy function is adopted as the loss function (Zhang and Sabuncu, 2018).

$$Loss = -\sum_{j=1}^{T} y_j log P_j \tag{3}$$

The data of 24 people in the THU benchmark dataset is used as the pre-training dataset, and different combinations of multiple people are randomly used for feature fusion. After repeated experiments to verify,the different combinations of multiple people used for feature fusion during the pre-training of the initialized feature extractor do not have an impact on the final classifier. An initial pretraining learning rate of 0.001 was adopted. The epoch was set to 80 and the mini-batch size of stochastic gradient descent is set to 16. Next, the pre-training model with initial weights was established for the experimental paradigm followed in this study. Based on the pre-training model, the epoch was then reset as 40 for training with the collected experimental data. A very large epoch makes a personal private network and reduces the generalization ability (Pan and Yang, 2009).

Results

The 10 subjects were labeled as S1 to S10, and two adjacent subjects made up one group (e.g., S1 and S2 made up group C1, and S3 and S4 made up C2). **Table 4** shows only one grouping case to show the fusion of features at different levels of feature quality, group members were interchangeable and tried different combinations. As the parallel feature connection method gives the best classification result with fused features and CNN, **Table 4** only lists the classification of different subjects and subject groups in different time windows under parallel feature connection. The three-channel EEG data of the two people in each group were connected in parallel to obtain six-channel EEG data, and the six-channel CNN model was trained using the TL strategy. **Table 4** shows that, based on the CNN classifier, the classification accuracy of both single- and two-person feature fusion declines as the time window decreases. Personal characteristics become more marked when the stimulation time is longer. This table compares the classification accuracy results of single-person CNN and five representative results of two grouping types with two people in a group: (1) Feature fusion of subject data with significant and insignificant features. That is, one of them yielded a high classification accuracy, but the other yielded a low classification accuracy. The final result is lower than the best single-person classification accuracy with significant features. (2) Feature fusion of subjects' data with only significant features. As the data features of the two people were both significant, the classification accuracy of the neural network was markedly improved.

Taking the 3-s time window as an example, the 10-person average classification accuracy of the single-person system CNN without TL is only 43.5%, but with TL, it can reach as high as 90.6%. The five-person average classification accuracy of the two-person CNN without TL is only 55.0%, while with TL, it reaches 96.6%.

The 10-person data containing S1–S10 were used as data sets for subsequent experiments and called the collected data set. The results of each model training and prediction are different. The collected data are randomly shuffled, and then feature fusion is performed to calculate the average classification accuracy and ITR through the 10-fold cross-validation method, as shown in **Figure 5**, respectively. It can be clearly seen from **Figure 5** that the classification accuracy results of the two feature fusion methods and feature averaging method based on CNN invariably exceed that of single-person CNN in different time windows. Three multi-person fusion methods based on CNN ITR significantly outperformed single-person CNN ($p < 0.0001$). Parallel feature connection ITR

TABLE 4 Classification accuracy comparison of single- and two-person models under different time windows.

	TL-CNN (%)							
	3 s	**2.8 s**	**2.6 s**	**2.4 s**	**2.2 s**	**2.0 s**	**1.8 s**	**1.6 s**
S1	91.6	87.5	81.3	77.1	79.1	75	70.8	60.4
S2	89.5	85.4	87.5	83.2	77.1	66.7	66.7	66.6
S3	97.9	91.6	91.6	89.6	91.7	83.3	77.1	70.8
S4	79.1	68.6	77	70.8	66.7	68.6	70.8	64.5
S5	100	100	97.9	95.8	95.8	93.7	87.5	75
S6	93.7	87.5	83.3	83.3	83.4	79.2	77.1	72.9
S7	79.1	77	77.1	68.8	60.4	60.4	58.3	54.2
S8	100	100	100	97.9	97.9	93.4	87.5	85.4
S9	75	64.5	68	70.9	60.4	60.4	60.4	50
S10	100	97.9	97.9	93.7	91.7	91.7	72.9	72.9
Saverage	90.6	86.4	86.2	82.4	80.4	77.3	72.9	67.3
C1	95.8	93.3	87.5	79.2	91.6	77	75.1	75
C2	89.5	93.3	95.5	85.4	81.2	79.2	81.3	75
C3	100	95.8	97.9	95.8	91.7	85.4	91.2	91.2
C4	100	100	95.8	93.7	87.5	85.4	85.4	75
C5	97.9	97.9	93.7	93.7	83.3	87.5	81.2	70.8
Coverage	96.6	96.1	94.1	89.6	87.1	83.1	82.9	77.4

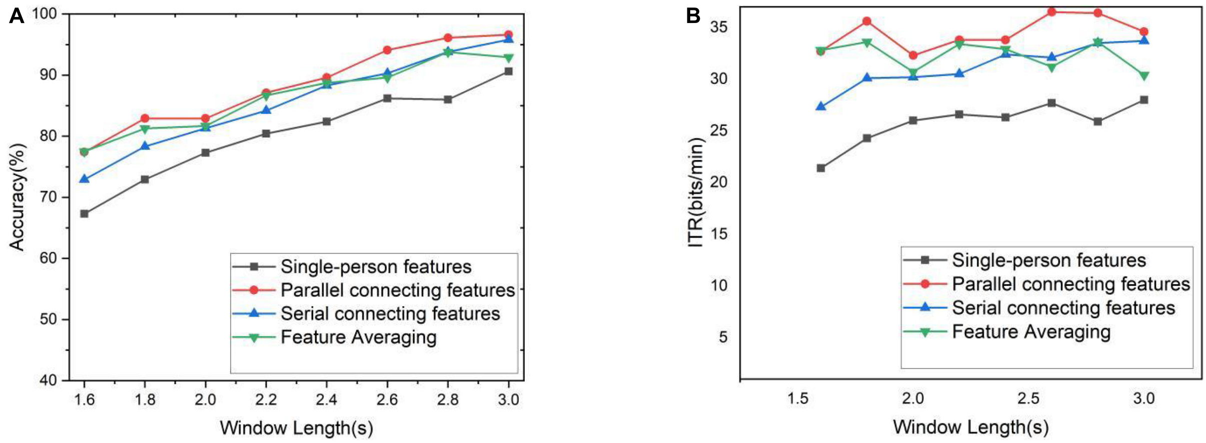


FIGURE 5
Accuracy and ITR under different time windows. **(A)** Classification accuracy for different time windows. **(B)** ITR for different time windows.

also significantly outperformed the other two feature fusion methods ($p < 0.05$). Among these three methods, the parallel feature connection method always ranks first, with the highest classification accuracy and ITR. The serial feature concatenation method and the feature averaging method exhibit similar overall performance, but the feature averaging method is more flexible and requires less computation. It can thus be concluded that feature averaging is better than serial feature concatenation.

The optimal scheme to set up cBCI is to train the corresponding parallel feature connection model with the TL method in advance when the number of subjects is known or to train the single-person model with the TL method and then apply the feature averaging method to this trained model when the number of subjects is unknown.

The classification accuracy of different time windows was averaged to obtain the total classification accuracy as shown in **Figure 6**. The total classification accuracy of the single-person CNN is 80.4% as baseline, which is far lower than the total classification accuracy of the multi-person CNN-based three feature fusion methods. As can be seen from **Figure 6**, when the number of participants in the experiment increased, the total classification accuracy of the three feature

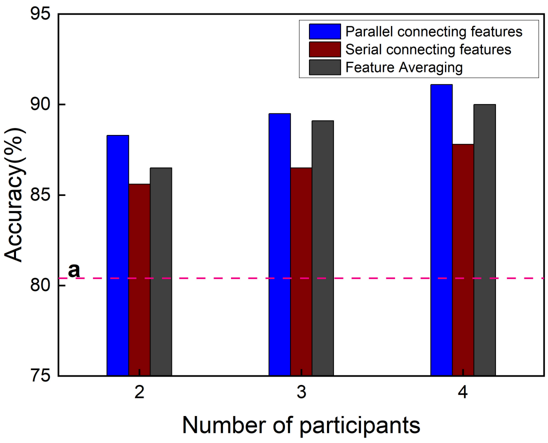


FIGURE 6
Total classification accuracy of three fusion methods under different number of participants. The line "a" is the total classification accuracy of the single features.

fusion methods slightly improved. The fusion method of parallel feature connection invariably attained the highest total classification accuracy; the feature averaging method was always noted to be the second best, and the serial feature connection was found to be the worst. Why is the parallel method so good? Upon increasing the number of participants, owing to the increase in the number of features, the CNN used by the three feature fusion methods needs to be slightly modified, mainly by increasing the number of convolution kernels. However, as the parallel feature connection needs to continuously integrate the information of multiple-lead channels through convolution, more convolutional layers are added. It has been proven that the convolution method can integrate the features of multiple individuals and multiple leads in a nonlinear way, which is better than the method of feature averaging. Therefore, if the computing power of the computer allows, it is an excellent cBCI construction method to use the method of parallel feature connection to fuse the EEG features of multiple people and send them into the TL-CNN model.

Discussion

Visually evoked potentials have been extensively studied by researchers (Mary Judith and Baghavathi Priya, 2021). When subjects gaze at flickering visual stimuli with flicker frequencies exceeding 4 Hz, their cerebral cortical activity would be modulated, resulting in a periodic rhythm similar to that of the stimulus (Bondre and Kapgate, 2014). CNN-based EEGNet, which was specially designed for EEG, has been widely applied to classification tasks in various EEG paradigms: e.g., P300 VEP, falsely correlated negatively matched waves, motor-related cortical potentials, and sensorimotor rhythms. In this study, the original EEGNet was modified, and the accuracy of the classification mode was improved by increasing the number of convolutional layers and the number size of convolutional kernel of each convolutional layer. In addition, three different multi-person EEG feature fusion methods are proposed herein to integrate multi-person EEG information to improve BCI performance. Each of the three fusion methods was found to have merits and shortcomings. In summary, in the case of known multi-person BCI collaborations with a fixed number of people, the parallel feature connection method is the best choice because it involves smaller models and fewer training parameters than the serial connection method; also, compared with feature averaging method, it produces higher classification accuracy. When the number of subjects is unknown, the feature averaging method should be chosen, and feature averaging using multiple subjects can be directly applied to a trained single-person CNN. Serial feature concatenation method is not suitable for the construction of online BCI system.

In this study, a small number of leads (e.g., O1, OZ, O2) were collected for setting up a multiple-person BCI system, which can facilitate the experimental preparation, reduce the subjects' fatigue, and improve the system's execution efficiency. Different from the voting and averaging methods of the existing distributed multi-person BCI systems, a complete single EEG data is constructed through multi-person feature fusion, and a transfer learning-based CNN is used to achieve classification in this new system. Compared with traditional CNN methods, the number of samples to collect is markedly reduced, and the accuracy is slightly improved. Moreover, a neural network with multiple narrow-band convolution kernels is constructed, and a multi-channel and multi-person feature fusion method is set up to extract the corresponding nonlinear features for fusion so as to improve the recognition accuracy of SSVEP-cBCI, further enhancing the classification accuracy and signal-to-noise ratio. The experimental results of the 10 subjects show that the CNN classification fusing two persons' features produces a higher SSVEP-cBCI recognition accuracy, and the TL-CNN-based two-person BCI effectively raises the classification accuracy and the robustness of BCI. The impact of individual differences in single-person BCI systems on system performance stability has been resolved. With the increase in the number of participants in the experiment, the total classification accuracy of the three feature fusion methods has been slightly improved, and the parallel feature connection method invariably exhibits the best performance.

The deep learning SSVEP-cBCI algorithm based on multi-person feature fusion established in this paper has been verified through offline system experiments and can be extended to real-time online systems in the future to complete real-time control of external equipment. Since the SSVEP-cBCI experimental paradigm requires multiple subjects to simultaneously fixate on the same flickering stimulus interface, a method of replicating multiple monitors was employed, and the collected multi-person EEG data were used for subsequent data processing and identification by a microcomputer. In the case of a fixed number of multi-person BCI collaborations and the computing power is allowed, it is necessary to prepare multiple corresponding number of different participants CNN classification models, perform corresponding feature fusion (parallel feature connection) and classification model training, and use the trained models to complete real-time online experiments. This feature fusion method can maximize the performance of cBCI. If the number of multi-person BCI collaborations is not fixed, or the computing power is not allowed, or there is not enough corresponding CNN classification model established, then the multi-person features can be integrated by the method of feature averaging, and the single-person CNN model can be used to complete real-time online experiments. Parallel feature connection are suitable for high-precision tasks, such as controlling unmanned vehicles, which requires precise control of the vehicle's travel to avoid collision. The feature averaging method fits into rehabilitation centers. Different numbers of patients can send requests at the same time, and as the number increases, medical staff can make corresponding responses more accurately. Serial feature connection can be used as an alternative method to increase the robustness of cBCI.

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 Ethics Committee of Tianjin Union Medical Center (the ethics agreement document number is TJUMC-2022-C05). The 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

PL: conceptualization and supervision. JS: data curation, software, and formal analysis. LC: methodology and validation. PL and JS: resources writing—original draft preparation. CC and AB: validation and writing—review and editing. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the National Natural Science Foundation of China (61806146) and the Natural Science Foundation of Tianjin City (18JCYBJC95400 and 19JCTPJC56000).

Conflict of interest

LC was employed by the China Electronics Cloud Brain (Tianjin) Technology Co., Ltd.

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

Publisher's note

References

Agarap, A. F. (2018). Deep learning using rectified linear units (relu). *arXiv* [Preprint] arXiv:1803.08375, doi: 10.48550/arXiv.1803.08375

Ali, P. J. M., Faraj, R. H., Koya, E., Ali, P. J. M., and Faraj, R. H. (2014). Data normalization and standardization: a tec'hnical report. *Mach. Learn. Tech. Rep.* 1, 1–6. doi: 10.13140/RG.2.2.28948.04489

Angrisani, L., Arpaia, P., Moccaldi, N., and Esposito, A. (2018). "Wearable augmented reality and brain computer interface to improve human-robot interactions in smart industry: a feasibility study for ssvep signals," in *Proceedings 2018 IEEE 4th International Forum on Research and Technology for Society and Industry (RTSI)*, (Piscataway, NJ: IEEE), 1–5. doi: 10.1109/RTSI.2018.8548517

Ba, J. L., Kiros, J. R., and Hinton, G. E. (2016). Layer normalization. *arXiv* [Preprint] arXiv:1607.06450, doi: 10.48550/arXiv.1607.06450

Bianchi, L., Gambardella, F., Liti, C., and Piccialli, V. (2019). "Group study via collaborative BCI," in *Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, (Piscataway, NJ: IEEE), 272–276. doi: 10.1109/SMC.2019.8914482

Bondre, C., and Kapgate, D. (2014). Brain computer interfaces using SSVEP: an overview. *Int. J. Eng. Technol.* 1, 9–17.

Cecotti, H. (2011). A time–frequency convolutional neural network for the offline classification of steady-state visual evoked potential responses. *Pattern Recog. Lett.* 32, 1145–1153. doi: 10.1016/j.patrec.2011.02.022

Cecotti, H., and Graser, A. (2010). Convolutional neural networks for P300 detection with application to brain-computer interfaces. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 433–445. doi: 10.1109/TPAMI.2010.125

Chai, X., Zhang, Z., Guan, K., Lu, Y., Liu, G., Zhang, T., et al. (2020). A hybrid bci-controlled smart home system combining ssvep and emg for individuals with paralysis. *Biomed. Signal Process. Control* 56:101687. doi: 10.1016/j.bspc.2019.101687

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

Dehzangi, O., and Farooq, M. (2018). Portable brain-computer interface for the intensive care unit patient communication using subject-dependent SSVEP identification. *BioMed Res. Int.* 2018:9796238. doi: 10.1155/2018/9796238

Dwarampudi, M., and Reddy, N. V. (2019). Effects of padding on LSTMs and CNNs. *arXiv* [Preprint] arXiv:1903.07288, doi: 10.48550/arXiv.1903.07288

Farmaki, C., Krana, M., Pediaditis, M., Spanakis, E., and Sakkalis, V. (2019). "Single-channel SSVEP-based BCI for robotic car navigation in real world conditions," in *Proceedings of the 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*, (Piscataway, NJ: IEEE), 638–643. doi: 10.1109/BIBE.2019.00120

Jang, E., Gu, S., and Poole, B. (2016). Categorical reparameterization with gumbel-softmax. *arXiv* [Preprint] arXiv:1611.01144, doi: 10.48550/arXiv.1611.01144

Katzenbach, J. R., and Smith, D. K. (2015). *The Wisdom Of Teams: Creating The High-Performance organization*. Boston, MA: Harvard Business Review Press.

Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., and Lance, B. J. (2018). EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *J. Neural Eng.* 15:056013. doi: 10.1088/1741-2552/aace8c

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324. doi: 10.1109/5.726791

Li, Q., Lu, Z., Gao, N., and Yang, J. (2019). Optimizing the performance of the visual P300-speller through active mental tasks based on color distinction and modulation of task difficulty. *Front. Hum. Neurosci.* 13:130. doi: 10.3389/fnhum.2019.00130

Mary Judith, A., and Baghavathi Priya, S. (2021). Multiset task related component analysis (M-TRCA) for SSVEP frequency recognition in BCI. *J. Ambient Intell. Hum. Comput.* 12, 5117–5126. doi: 10.1007/s12652-020-01962-8

Munzert, J., Lorey, B., and Zentgraf, K. (2009). Cognitive motor processes: the role of motor imagery in the study of motor representations. *Brain Res. Rev.* 60, 306–326. doi: 10.1016/j.brainresrev.2008.12.024

Nayak, T., Ko, L. W., Jung, T. P., and Huang, Y. (2019). "Target classification in a novel SSVEP-RSVP based BCI gaming system," in *Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC) (pp. 4194-4198)*, (Piscataway, NJ: IEEE), doi: 10.1109/SMC.2019.8914174

Pan, S. J., and Yang, Q. (2009). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359. doi: 10.1109/TKDE.2009.191

Peng, Y., Wong, C. M., Wang, Z., Wan, F., Vai, M. I., Mak, P. U., et al. (2019). Fatigue evaluation using multi-scale entropy of EEG in SSVEP-based BCI. *IEEE Access* 7, 108200–108210. doi: 10.1109/ACCESS.2019.2932503

Ravi, A., Beni, N. H., Manuel, J., and Jiang, N. (2020). Comparing user-dependent and user-independent training of CNN for SSVEP BCI. *J. Neural Eng.* 17:026028. doi: 10.1088/1741-2552/ab6a67

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. Healthcare Eng.* 2020:6968713. doi: 10.1155/2020/6968713

Song, X., Zeng, Y., Tong, L., Shu, J., Yang, Q., Kou, J., et al. (2022). A collaborative brain-computer interface framework for enhancing group detection performance of dynamic visual targets. *Comput. Intell. Neurosci.* 2022:4752450. doi: 10.1155/2022/4752450

Vaid, S., Singh, P., and Kaur, C. (2015). "EEG signal analysis for BCI interface: A review," in *Proceedings 2015 fifth International Conference On Advanced Computing & Communication Technologies*, (Piscataway, NJ: IEEE), 143–147. doi: 10.1109/ACCT.2015.72

Valeriani, D., Poli, R., and Cinel, C. (2016). Enhancement of group perception via a collaborative brain–computer interface. *IEEE Trans. Biomed. Eng.* 64, 1238–1248. doi: 10.1109/TBME.2016.2598875

Wang, Y., and Jung, T. P. (2011). A collaborative brain-computer interface for improving human performance. *PLoS One* 6:e20422. doi: 10.1371/journal.pone.0020422

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.

Zhang, X., Xu, G., Mou, X., Ravi, A., Li, M., Wang, Y., et al. (2019). A convolutional neural network for the detection of asynchronous steady state motion visual evoked potential. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 1303–1311. doi: 10.1109/TNSRE.2019.2914904

Zhang, Z and Sabuncu, M. R. (2018). Generalized cross entropy loss for training deep neural networks with noisy labels. In Proceedings of the 32nd International Conference on Neural Information Processing Systems. pp. 8792–8802. doi: 10.48550/arXiv.1805.07836

Zhao, X., Chu, Y., Han, J., and Zhang, Z. (2016). SSVEP-based brain–computer interface controlled functional electrical stimulation system for upper extremity rehabilitation. *IEEE Trans. Syst. Man Cybern.* 46, 947–956. doi: 10.1109/TSMC.2016.2523762

frontiers | Frontiers in Neuroscience

TYPE Correction
PUBLISHED 26 October 2022
DOI 10.3389/fnins.2022.1024150

Corrigendum: Multi-person feature fusion transfer learning-based convolutional neural network for SSVEP-based collaborative BCI

Penghai Li[1], Jianxian Su[1], Abdelkader Nasreddine Belkacem[2], Longlong Cheng[3]* and Chao Chen[4]*

[1]School of Integrated Circuit Science and Engineering, Tianjin University of Technology, Tianjin, China, [2]Department of Computer and Network Engineering, College of Information Technology, UAE University, Al Ain, United Arab Emirates, [3]China Electronics Cloud Brain Technology Co., Ltd., Tianjin, China, [4]Key Laboratory of Complex System Control Theory and Application, Tianjin University of Technology, Tianjin, China

OPEN ACCESS

EDITED AND REVIEWED BY
Michele Giugliano,
International School for Advanced Studies (SISSA), Italy

*CORRESPONDENCE
Longlong Cheng
chenglonglong@cecdat.com
Chao Chen
cccovb@hotmail.com

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

RECEIVED 21 August 2022
ACCEPTED 12 October 2022
PUBLISHED 26 October 2022

CITATION
Li P, Su J, Belkacem AN, Cheng L and Chen C (2022) Corrigendum: Multi-person feature fusion transfer learning-based convolutional neural network for SSVEP-based collaborative BCI. *Front. Neurosci.* 16:1024150. doi: 10.3389/fnins.2022.1024150

KEYWORDS

steady-state visually evoked potential, collaborative BCI, feature fusion, convolutional neural network, transfer learning

A corrigendum on

Multi-person feature fusion transfer learning-based convolutional neural network for SSVEP-based collaborative BCI

by Li, P., Su, J., Belkacem, A. N., Cheng, L., and Chen, C. (2022). *Front. Neurosci*. 16:971039. doi: 10.3389/fnins.2022.971039

In the published article, there was an error in the article title as published. Instead of "Steady-state visually evoked potential collaborative BCI system deep learning classification algorithm based on multi-person feature fusion transfer learning-based convolutional neural network," it should be "Multi-person feature fusion transfer learning-based convolutional neural network for SSVEP-based collaborative BCI."

The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

Publisher's note

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 05 August 2022
DOI 10.3389/fnins.2022.972581

OPEN ACCESS

EDITED BY
Hui Zhou,
Nanjing University of Science and Technology, China

REVIEWED BY
Shixiong Chen,
Shenzhen Institutes of Advanced Technology (CAS), China
Ziliang Xu,
Fourth Military Medical University, China

*CORRESPONDENCE
Weiping Zheng
zhengweiping@scnu.edu.cn

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

RECEIVED 18 June 2022
ACCEPTED 18 July 2022
PUBLISHED 05 August 2022

CITATION
Chen J, Shen M, Ma W and Zheng W (2022) A spatio-temporal learning-based model for sleep apnea detection using single-lead ECG signals. *Front. Neurosci.* 16:972581. doi: 10.3389/fnins.2022.972581

A spatio-temporal learning-based model for sleep apnea detection using single-lead ECG signals

Junyang Chen[1], Mengqi Shen[2], Wenjun Ma[1] and Weiping Zheng[1]*

[1]School of Computer Science, South China Normal University, Guangzhou, China, [2]Grado Department of Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States

Sleep apnea (SA) is a common chronic sleep breathing disorder, which would cause stroke, cognitive decline, cardiovascular disease, or even death. The SA symptoms often manifest as frequent breathing interruptions during sleep and most individuals with sleeping disorders are not aware of the SA events. Using a portable device with single-lead ECG signal is an effective way to help an individual to monitor their sleep conditions at home. However, the SA detection performance of ECG-based methods is still difficult to meet the clinical practice requirement. In this study, we propose an end-to-end spatio-temporal learning-based SA detection method, which consists of multiple spatio-temporal blocks. Each block has the identical architecture with a convolutional neural network (CNN) layer, a max-pooling layer, and a bi-gated recurrent unit (BiGRU) layer. This architecture with repeated spatio-temporal blocks can well capture the morphological spatial feature information as well as the temporal feature information from ECG signals. The proposed SA detection model was evaluated on the publicly available datasets of PhysioNet Apnea-ECG dataset (Apnea-ECG) and University College Dublin Sleep Apnea Database (UCDDB). Extensive experimental results show that our proposed SA model on both Apnea-ECG and UCDDB datasets achieves state-of-the-art results, which are obviously superior to existing ECG-based SA detection methods. It means that our proposed method has the potential to be deployed into a healthcare system to provide a sleep monitoring service, which can screen out SA population with high risk and help to take timely interventions to prevent serious consequences.

KEYWORDS

sleep apnea, ECG signals, spatio-temporal learning, BiGRU, attention

1. Introduction

Sleep apnea (SA) is a sleep disorder in which breathing is interrupted several times during sleeping. Its typical symptoms include headache, insomnia and others, and it can be potentially serious (Russell et al., 2014). Without prompt and appropriate treatment measures, patients with SA would suffer from serious complications such as stroke

(King and Cuellar, 2016), cognitive decline (Vanek et al., 2020), cardiovascular disease (Lin et al., 2020), and even death. SA is considered by some researchers to be a recognized independent risk factor for stroke, such that individuals with SA have an approximately twofold greater risk of stroke compared with those without SA (Lyons and Ryan, 2015). This indicates that SA is a great threat to the global physical and mental health, with ∼936 million adults (male and female) aged 30–69 years worldwide suffering from mild to severe obstructive sleep apnea, and 425 million adults aged 30–69 years suffering from moderate to severe obstructive sleep apnea (Benjafield et al., 2019). Due to the prevalence of SA, it is very vital to screen out individuals with SA and take timely interventions.

In clinical practice, polysomnography (PSG) is the gold standard test to diagnose SA. However, conducting PSG is expensive and often unavailable due to the shortage of physical therapists and sleep monitoring units (Graco et al., 2018). PSG with many biomedical sensors, including electroencephalogram (EEG), electro-oculogram (EOG), electromyogram (EMG), electrocardiogram (ECG) and pulse oximetry as well as airflow and respiratory effort, is performed as an SA test overnight in a sleep laboratory or specific unit in a hospital (Ali et al., 2019). This can be quite cumbersome and uncomfortable, so the collected signals cannot reflect the individuals' sleep conditions. In addition, a physical therapist must be available when the PSG is conducted in the hospital, which significantly restricts the screening of people with SA.

Home Sleep Test (HST) is an alternative to PSG for SA diagnosis, which is usually conducted overnight outside of the hospital or sleep laboratory (Rosen et al., 2017). Portable devices, which are simple, of low-cost and easy to operate, have been developed to enable the patients to monitor their sleep conditions in an unattended home environment. Gaiduk et al. (2020) have developed a system based on pressure sensors that can work independently and *via* wireless connection, which is as accurate as the current technology. However, this pressure sensor-based approach is highly sensitive to pressure, and the pressure signal can be easily contaminated with noise from the external environment. ECG is considered to be one of the most relevant physiological signals for the SA detection because patient's heart rate increases when SA occurs (Somers et al., 2008; Wang T. et al., 2019). ECG contains valuable information about the cardiorespiratory system and is therefore of great importance for SA detection (Bahrami and Forouzanfar, 2021). Over the past twenty years, various approaches have been proposed for the automated SA detection using HRV and EDR signals which can be derived from ECG (Gutiérrez-Tobal et al., 2015; Faust et al., 2016; Smruthy and Suchetha, 2017; Viswabhargav et al., 2019). In addition, ECG can be easily recorded. Therefore, ECG-based portable devices represent a better option. A wearable ECG acquisition system has been developed (Weder et al., 2015), which is designed as a chest strap that can continuously monitor ECG signals for multiple nights. As a more comfortable option, Ankhili et al. (2018) developed a reliable, washable ECG monitoring undergarment that can record and send the ECG signal wirelessly to a smartphone to analyze the ECG signal in real-time.

Using ECG signals can greatly reduce the complexity of diagnostic SA tests and allow for better monitoring of physiological changes in the patient (Bsoul et al., 2011). Several algorithms have been proposed for ECG-based SA detection. These algorithms generally include a first step of feature extraction from the original ECG signals, and then these features are used as the input and fed into a classification model (Baty et al., 2020). Sharma and Sharma (2016) extracted features from QRS complex waves using Hermite decomposition. Then, these features were combined with time series features, and least squares-support vector machine (LS-SVM) was used as a classifier for SA detection. Song et al. (2015) introduced a classifier that blends an SVM with a hidden Markov model (HMM) to take advantage of the time-dependent nature of SA segments. In recent years, deep neural networks (DNN) with end-to-end training are also applied to build SA detection models. Li et al. (2018) used HMM, ANN, and ECG signals for the identification of SA segments. Feng et al. (2020) used an unsupervised neural network to learn features, and they improved the performance of the classifier by taking into account the time-dependence and imbalance problems.

Although aforementioned models achieved promising results, there still exists a gap between their SA detection performance and the requirement of actual applications. Note that spatial patterns and temporal correlations are both important for SA detection. For example, the R-peak has salient spatial features on the ECG signals; while the RR intervals can be derived from the temporal dependencies. In reality, RR intervals are frequently utilized in SA detection (Bahrami and Forouzanfar, 2021). However, the spatio-temporal correlations are seldom utilized in the existing SA detection models (Sharan et al., 2020; Chen et al., 2021; Yang et al., 2022). Bahrami and Forouzanfar (2021) has used a hybrid CNN and LSTM network to extract spatio-temporal feature from ECG signals. However, they only use a simple combination of CNN and LSTM networks. To improve the performance of SA detection, a spatio-temporal learning based DNN model is proposed in this paper. In order to take advantage of more spatio-temporal dependencies, multiple adjacent segments are concatenated and used as the input of the proposed model. A spatio-temporal learning block is designed which is composed of a one-dimensional convolutional neural network (CNN), a max-pooling layer, and a bidirectional gated recurrent unit (BiGRU). Multiple spatio-temporal blocks are stacked in the proposed model to extract long-range spatial and temporal correlations. As a result, this model can fully utilize the multiple concatenated ECG segments. Moreover, an attention mechanism is employed to further utilize the global correlations by using the high-level forward and backward spatio-temporal features. These

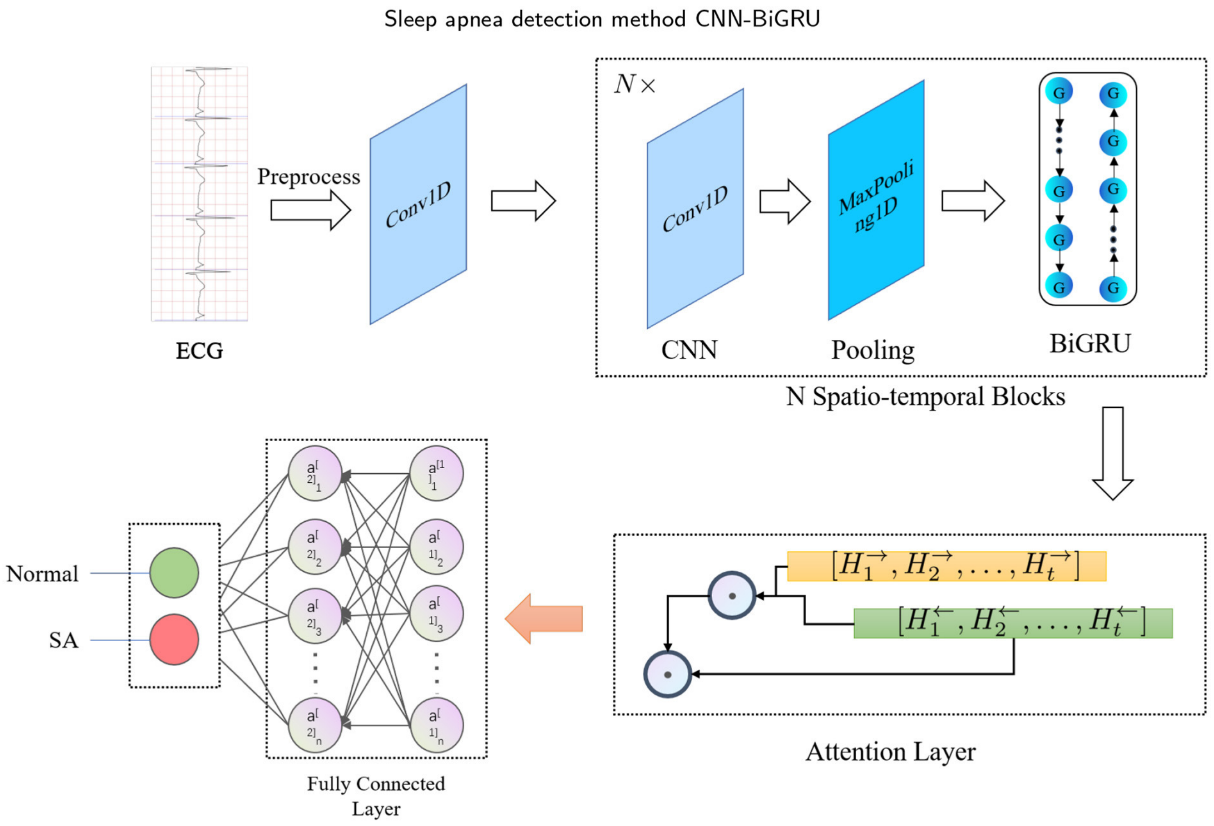


FIGURE 1
Flowchart of the proposed sleep apnea detection model.

characteristics have made our model different with the other spatio-temporal model (Bahrami and Forouzanfar, 2021). The performance of our model has outperformed the one of Bahrami and Forouzanfar (2021). Experimental results on two public domain datasets of PhysioNet Apnea-ECG dataset (Apnea-ECG) and University College Dublin Sleep Apnea Database (UCDDB) showed that CNN-BiGRU achieved the competitive performance to the previous state-of-the-art methods. The main contributions of this study can be listed as follows:

- To fully extract spatio-temporal information from ECG signals, we proposed a spatio-temporal learning-based model called CNN-BiGRU with multiple spatio-temporal blocks, each block of which consists of a one-dimensional CNN layer, a max-pooling layer, and a BiGRU layer.
- We employed an attention mechanism to further exploit the high-level forward and backward spatio-temporal features from the last spatio-temporal block, which was able to extract the global correlations from multiple ECG signal segments.
- Experiment results on two public domain datasets of Apnea-ECG and UCDDB showed that the proposed CNN-BiGRU achieved a state-of-the-art performance, which outperforms the previous state-of-the-art methods. It means that the proposed CNN-BiGRU can be potentially deployed into a medical system to provide the SA monitoring service.

The remainder of this paper is organized as follows. Section 2 details the composition of the CNN-BiGRU model. In Sections 3 and 4 the results are presented and discussed. The conclusion is presented in Section 5.

2. Methods

2.1. Overview

The main idea of this study is to develop a fully automated (or end-to-end) spatio-temporal learning-based SA detection method, which is illustrated in Figure 1. First, in the pre-processing phase, the RR intervals and R-peak amplitudes are extracted from combining adjacent and labeled segments. Then, the RR intervals and R-peak amplitudes are fed into the

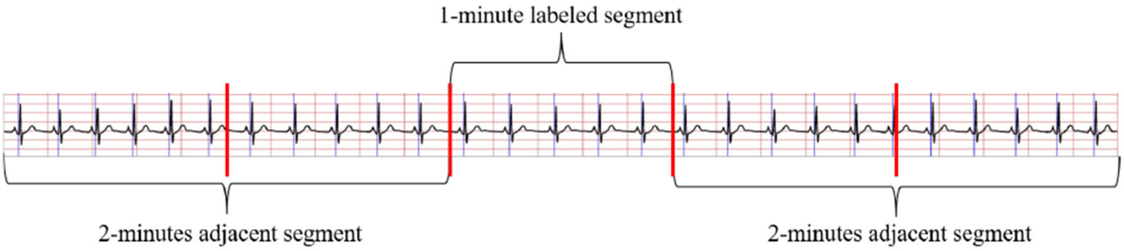


FIGURE 2
Five segments schematic diagram.

proposed CNN-BiGRU model, which recycles through multiple spatio-temporal blocks to capture high-level spatio-temporal features. The whole method can be mathematically defined as follows: Given the input $X \in \mathbb{R}^D$, a mapping function $f : X \rightarrow \{0, 1\}$ is learned, where D is the dimensionality of the input, and 0 and 1 denote normal and SA, respectively. Specifically, the spatio-temporal block consists of CNN, max-pooling and BiGRU layers. Furthermore, the attention mechanism is used to extract the most effective part of the spatio-temporal features and improve the accuracy. Finally, the fully connected layer is used to identify whether the labeled segment belongs to SA.

2.2. Preprocessing

The method in Wang T. et al. (2019) to obtain the RR intervals and R-peak amplitudes were applied in this study to pre-process ECG signals. Considering adjacent segment information is useful for SA detection in each segment. As shown in Figure 2, both the labeled segment and its surrounding ECG signal ±2 segments (five segments total of 1 min) were extracted and processed. Firstly, the Hamilton algorithm (Hamilton, 2002) was used to find R-peaks and adjust the detection peak to match the local signal maximum. Then, RR intervals were calculated and the amplitudes were extracted using the locations of the R-peaks, while anomalous signals were removed. For physiologically unexplained points, median filters were chosen to solve the extracted RR intervals (Chen et al., 2015). The final problem of unequal time intervals between the RR intervals and amplitudes was tackled by cubic interpolation, which yielded 900 RR interval points and 900 amplitudes over 5 min segments.

2.3. The proposed CNN-BiGRU for SA recognition

The proposed CNN-BiGRU consists of spatio-temporal blocks of CNNs and BiGRUs, an attention layer and fully connections layers. These layers are introduced in detail as follows.

2.3.1. Convolutional neural network

Convolutional neural networks are among the most common and efficient techniques that are widely used in various signal and image processing applications (Fan et al., 2018, 2019; Wang et al., 2018). A lightweight CNN model can be trained on a mixed-scale graph in order to extract deep features for the detection of obstructive SA (Mashrur et al., 2021). In this study, we used a one-dimensional CNN to extract spatial dependencies from ECG signals, which is mathematically defined as follows:

$$[\mathbf{w} \circledast \mathbf{x}](i) = \sum_{u=0}^{L-1} w_u x_{i+u} \tag{1}$$

where $\mathbf{x}, \mathbf{w}, L$ are the input, filter, and filter size, respectively.

The next layer of the CNN is generally the pooling layer. The max-pooling layer can be used to reduce mean value shift errors caused by bad initialization of weights (Wang L. et al., 2019). In this paper, the max-pooling layer was used to decrease the computational effort and to mitigate the overfitting problem by selecting the maximum value of each feature.

2.3.2. Gated recurrent unit

Gated recurrent units (GRU) (Cho et al., 2014) represent a more advanced alternative to the simple recurrent neural network (RNN) and are more capable of learning long-term dependencies than vanilla RNN (Zhang et al., 2022). While both GRU and Long Short-Term Memory (LSTM) units have gating units that regulate the flow of information within the unit, GRUs do not have a separate memory unit, only update and reset gates. The j-th hidden unit of each mentioned gate is defined as follows:

Reset gates:

$$r_j = \sigma\left([\mathbf{W}_r\mathbf{x}]_j + \left[\mathbf{U}_r\mathbf{h}_{\langle t-1\rangle}\right]_j\right) \tag{2}$$

where σ and $[\cdot]_j$ are the logistic sigmoid function and the j-th element of a vector, respectively; t is the time step, $\mathbf{x}$ denotes the input, and $\mathbf{h}_{\langle t-1\rangle}$ represents the previous hidden state.

Update gates:

$$z_j = \sigma\left([\mathbf{W}_z\mathbf{x}]_j + \left[\mathbf{U}_z\mathbf{h}_{\langle t-1\rangle}\right]_j\right). \tag{3}$$

Output:

$$h_j^{\langle t\rangle} = z_j h_j^{\langle t-1\rangle} + \left(1 - z_j\right)\tilde{h}_j^{\langle t\rangle} \tag{4}$$

The weight matrices $\mathbf{W}_r$, $\mathbf{W}_z$, $\mathbf{W}_x$, $\mathbf{U}_r$, $\mathbf{U}_z$, and $\mathbf{U}_x$ are learned during training. The candidate hidden state $\tilde{h}_j^{\langle t\rangle}$ is computed as follows:

$$\tilde{h}_j^{\langle t\rangle} = \tanh\left([\mathbf{W}_x\mathbf{x}]_j + r_j\left[\mathbf{U}_x\mathbf{h}_{\langle t-1\rangle}\right]\right) \tag{5}$$

The GRU can be heavily dependent on the dataset and the corresponding task, and the Apnea-ECG dataset works better with GRU than with LSTM, since it has fewer parameters and faster training (Chung et al., 2014). In this work, a bidirectional GRU was used to capture richer temporal information. By recursively computing the hidden states H_t in the forward and backward directions, the forward sequence F and the backward sequence B were obtained, respectively. This can be mathematically defined as follows:

$$H_t^{\rightarrow} = [h_1^{\langle t\rangle}, h_2^{\langle t\rangle} \dots h_n^{\langle t\rangle}] \tag{6}$$

$$H_t^{\leftarrow} = [h_1^{'\langle t\rangle}, h_2^{'\langle t\rangle} \dots h_n^{'\langle t\rangle}] \tag{7}$$

$$F = [H_1^{\rightarrow}, H_2^{\rightarrow}, \dots, H_s^{\rightarrow}]^T \in \mathbb{R}^{s\times n} \tag{8}$$

$$B = [H_1^{\leftarrow}, H_2^{\leftarrow}, \dots, H_s^{\leftarrow}]^T \in \mathbb{R}^{s\times n} \tag{9}$$

where n denotes the number of GRUs and s represents the total number of the time step.

2.3.3. Attention mechanism

Dot-product attention (Luong et al., 2015) was used to extract the global correlation information from the input multiple ECG segments. Specifically, dot-production attention is applied on the forward sequence F and the backward sequence B of the BiGRU unit within the last spatio-temporal block. The attention score is calculated as follows:

$$\text{Attention}\,(F, B, B) = \text{softmax}\left(FB^T\right)B \tag{10}$$

Using the attention mechanism allows the model to pay more attention to specific high-level spatial-temporal dependency information, improving the accuracy of the model. Note that dot-product attention is fast and spatially efficient because it enables a highly optimized code for matrix multiplication (Vaswani et al., 2017).

2.3.4. Proposed CNN-BiGRU model

To better extract the spatio-temporal features of ECG signals, we have specially designed a SA detection model, named CNN-BiGRU. The proposed CNN-BiGRU model is mainly composed of a CNN layer, multiple stacked spatio-temporal learning blocks, an attention layer, and fully connected layers. First, a convolutional layer was used to extract the base features before using spatio-temporal blocks. A spatio-temporal learning block consists of a one-dimensional CNN, a max-pooling layer, and a BiGRU unit. The use of multiple spatio-temporal blocks enables the CNN-BiGRU model to extract high-level spatio-temporal features from the ECG signal. Specifically, our model is able to extract local spatial features of the R-peaks, as well as global temporal features of the heartbeat intervals. Then, the attention score of the fused high-level forward and backward features from the spatio-temporal blocks was calculated. This attention mechanism is able to further utilize the global spatio-temporal correlations from the multiple ECG segments. Finally, three dense layers were used for classification. Additionally, some of the layers were immediately followed by a dropout layer to mitigate the effects of overfitting. The mathematical expression of the whole model computation process is as follows: For the CNN input X, the output C is:

$$C = g(f(X; \mathbf{W})) \tag{11}$$

where g denotes the ReLU activation function $g(x) = \max(0, x)$ and $\mathbf{W}$ denotes the convolution kernel.

As previously mentioned, this study uses BiGRU with the matrices $\mathbf{W}_r$, $\mathbf{U}_r$, $\mathbf{W}_z$, $\mathbf{U}_z$, $\mathbf{W}_x$, and $\mathbf{U}_x$ as the parameters to be learned. After reducing the size of the feature map through the max-pooling layer, the output C of the CNN was fed into the BiGRU. The output of the spatio-temporal block can be mathematically represented as follows:

$$F = \varphi(C; \mathbf{W}_r^{\rightarrow}, \mathbf{U}_r^{\rightarrow}, \mathbf{W}_z^{\rightarrow}, \mathbf{U}_z^{\rightarrow}, \mathbf{W}_x^{\rightarrow}, \mathbf{U}_x^{\rightarrow}) \tag{12}$$

$$B = \varphi(C; \mathbf{W}_r^{\leftarrow}, \mathbf{U}_r^{\leftarrow}, \mathbf{W}_z^{\leftarrow}, \mathbf{U}_z^{\leftarrow}, \mathbf{W}_x^{\leftarrow}, \mathbf{U}_x^{\leftarrow}) \tag{13}$$

where F and B are the stacked hidden states in the forward and backward directions, respectively. If the next layer of the BiGRU was a CNN, then F and B were concatenated along the channel dimension, otherwise the attention score a was calculated as follows:

$$a = \text{softmax}\left(FB^T\right)B \tag{14}$$

Finally, the attention score was entered into the fully connected layer for classification, and the labeled segments were classified to be SA or normal. Table 1 has listed the architecture details of the proposed CNN-BiGRU model. Specifically, the architecture contains three spatio-temporal blocks (see the layers of 2–5, 7–10, and 12–15, respectively in Table 1). And the dropout ratios in Table 1 have all been set to 0.2.

Algorithm 1 CNN-BiGRU training.

```
Input: Given training set $[X, Y] = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ and validation set $[\tilde{X}, \tilde{Y}]$,
    the CNN-BiGRU model $f$ with initialized parameters $\theta_0$, epochs $T$, learning rate $\alpha$
Output: Trained model weight $\Theta$
Initialize the parameters of the Adam optimizer: the exponential decay rates of the first and second order
    moments are estimated as $\beta_1$ and $\beta_2$
$m_0 = 0, v_0 = 0$
for $t = 1$ to $T$ do
    Forward-propagation: $\hat{y} = f(X; \theta_t)$
    Compute loss error:
        $J(\theta) = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \ln(\hat{y}_i) + (1 - y_i)\ln(1 - \hat{y}_i)\right]$
    Compute the gradient of the current data:
        $g_t = \frac{\partial}{\partial \theta_t} J(\theta_t)$
    Update network parameters by Adam optimizer:
        $m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$
        $v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$
        $\theta_t = \theta_{t-1} - \alpha \frac{\sqrt{(1-\beta_2^t)}m_t}{(1-\beta_1^t)(\sqrt{v_t}+\epsilon)}$
    if $f(\tilde{X}; \theta_t)$ turns out better on the validation set than before then
        Save weight $\theta_t$ to $\Theta$
    end if
end for
```

2.4. Experimental settings

In order to enable an enhanced performance of the CNN-BiGRU model, the number of spatio-temporal blocks was tuned from 1 to 5. Adam optimizer (Kingma and Ba, 2017) and binary cross-entropy loss function were applied for parameter optimization. The learning rate and batch size were set to 0.001 and 128, respectively. The proposed model was trained for 40 epochs. In each training epoch, the model parameters were evaluated using the validation set and the best model parameters were saved to perform classification on the test set. Detailed training methods are described in Algorithm 1. Our model was implemented using the Tensorflow framework with a Tesla P100-PCIE-16GB graphics card.

TABLE 1 Detailed parameter settings for the CNN-BiGRU model.

Layer	Type	Number of filter/cell/unit	Kernel size	Activation function
1	Convolutional	128	3	ReLU
2	Convolutional	128	3	ReLU
3	Max-Pooling	–	3	–
4	Dropout	–	–	–
5	Bidirectional GRU	128	–	Tanh
6	Dropout	–	–	–
7	Convolutional	128	3	ReLU
8	Max-Pooling	–	3	–
9	Dropout	–	–	–
10	Bidirectional GRU	128	–	Tanh
11	Dropout	–	–	–
12	Convolutional	128	3	ReLU
13	Max-Pooling	–	3	–
14	Dropout	–	–	–
15	Bidirectional GRU	128	–	Tanh
16	Attention	–	–	–
17	Flatten	–	–	–
18	Dense	64	–	ReLU
19	Dropout	–	–	–
20	Dense	64	–	ReLU
21	Dense	2	–	Softmax

Various performance metrics, such as precision, specificity, F_1 score, sensitivity, and accuracy, were used to assess the performance of the proposed model. These metrics are defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{15}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{16}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{17}$$

$$F_1 \text{ score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{18}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{19}$$

where FP, TP, FN, and TN stand for "false positive," "true positive," "false negative," and "true negative," respectively. The SA class is the positive class in this study, while the normal class is the negative class.

This model was also evaluated using the receiver operating characteristic (ROC) and the related area under the curve (AUC).

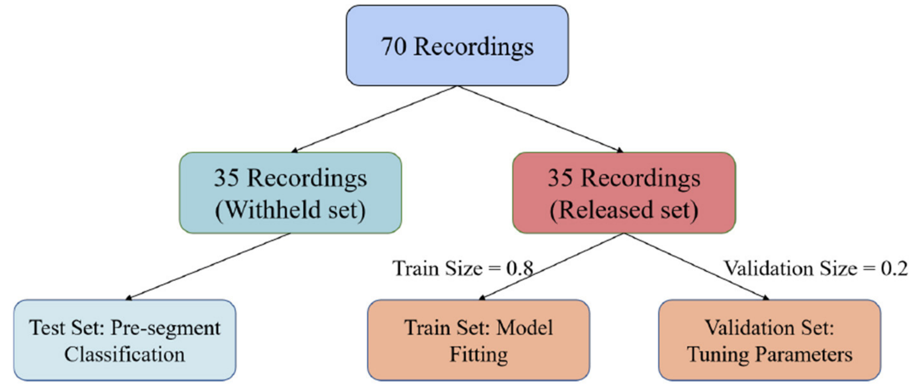


FIGURE 3
Datasets division methods on PhysioNet Apnea-ECG dataset.

3. Experimental results

3.1. Datasets

3.1.1. PhysioNet Apnea-ECG dataset

In this paper, we used the PhysioNet Apnea-ECG dataset provided by Philipps University (Penzel et al., 2000) for model evaluation. The Apnea-ECG dataset has 70 recordings, including 35 records in the released dataset (a01–a20, b01–b05, and c01–c10) and 35 records in the withheld dataset (x01–x35). And the release set is used to train the model with the withheld set used to test the model. Regarding the released set, 20% of the 35 released data were used to validate the model and tune its hyperparameters (Figure 3). ECG recordings for this dataset were obtained from subjects with an AHI (apnea hypoventilation index) between 0 and 83. And these subjects ranged in age from 27 to 63 years and their body mass indices varied between 19.2 and $45.33 kg/m^2$. The ECG signal was sampled at 100 Hz over a range of 401 to 587 min. The experts labeled each 1 min recording segments as SA or normal. According to our pre-proccessing method, the release and withheld set contained 17,045 and 17,302 segments, respectively. The experimental results show that CNN-BiGRU achieves an outstanding performance in SA detection.

3.1.2. University college Dublin sleep Apnea database (UCDDB)

UCDDB was used as a second dataset to validate the performance of CNN-BiGRU. This database contains 25 full overnight polysomnograms from adult subjects suspected sleep-disordered breathing. ECG recordings of this dataset have been collected by a modified lead V2. We used ECG signals sampled at 128 Hz and with the durations ranging from 355 to 462 min. Following previous studies (Mostafa et al., 2017, 2020), we labeled a 1 min segment as SA if the segment contains more than 5 s of SA events. Considering the class imbalance problem of UCDDB, the data of patients without SA events (ucddb008, ucddb011, ucddb013, and ucddb018) are not used (John et al., 2021).

3.2. Classification performance on Apnea-ECG dataset

The SA detection involves two stages. The first stage is to detect whether a 1 min segment is SA. In the second stage, each patient is assessed for sleep quality overnight based on the results of the first stage.

3.2.1. Per-segment classification on Apnea-ECG dataset

Test sets were used to evaluate the effectiveness of the proposed model. First, the pre-processed 5 min ECG segments were fed into the CNN-BiGRU network to automatically extract the features. Then, the extracted features were fed into the fully connected layers, and the ECG signal of the middle segment was classified. The results of the CNN-BiGRU model with three spatio-temporal blocks for ten runs are listed in the Table 2. It is worth noting that the 10th experiment exceeded the average on all evaluation metrics. In addition, the 5th experiment reached the highest values of 91.54 and 88.82% for the accuracy and F_1 score, respectively. To evaluate the classifier more comprehensively, Figure 4 shows the ROC curve and AUC. It can be seen that the model proposed in this study is stable, with an AUC value of 0.9692 ± 0.0013.

Table 3 lists the comparison results between the CNN-BiGRU model and previous state-of-the-art works on the detection of per-minute ECG signals. Notice that the compared methods and the proposed models listed in Tables 3, 4 were all trained on the release set and evaluated on the withheld set.

TABLE 2 Per-minute detection performance Results on the Apnea-ECG dataset.

	Accuracy (%)	Recall (%)	Specificity (%)	Precision (%)	F_1 score (%)
1	91.25	85.69	94.71	90.96	88.24
2	90.80	86.84	93.26	88.88	87.85
3	91.18	88.10	93.09	88.79	88.45
4	91.14	88.37	92.86	88.49	88.43
5	91.54	87.72	93.92	89.95	88.82
6	90.95	86.38	93.79	89.62	87.97
7	91.18	84.84	95.12	91.52	88.05
8	91.36	83.78	96.07	92.97	88.13
9	91.37	86.41	94.45	90.63	88.47
10	91.41	86.64	94.38	90.53	88.54
Mean	91.22	86.48	94.16	90.23	88.30
Std	0.2098	1.360	0.9404	1.316	0.2833

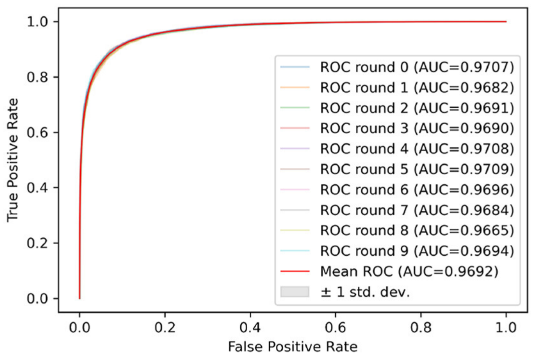


FIGURE 4
ROC curves for 10 random repeated runs on PhysioNet Apnea-ECG dataset.

The results show that the average performance of CNN-BiGRU outperforms the previous optimal model in terms of accuracy, specificity, and F_1 metrics. It is worth noting that CNN-BiGRU only underperforms the approach of literature (Li et al., 2018; Yang et al., 2022) in terms of the recall metric. However, our model achieved the F_1 score of 88.3%, which is better than that of literature (Yang et al., 2022), while literature (Li et al., 2018) did not give F_1 score.

In summary, some previous works (Song et al., 2015; Sharma and Sharma, 2016) were based on feature engineering techniques that attempt to improve the performance by mapping high-dimensional training data to a low-dimensional feature space. However, this is time-consuming and ineffective. Deep learning methods can extract important features from ECG signals, and the DL-based methods (Chen et al., 2021; Yang et al., 2022) mentioned in Table 3 have all achieved good results, but their performance was inferior to that of the CNN-BiGRU model proposed in this paper. Our model uses spatio-temporal blocks, which can extract spatio-temporal features more effectively from ECG signals and provide better performance on SA classification.

3.2.2. Per-recording classification on Apnea-ECG dataset

In order to further assess the quality of the subjects' sleep, an overall SA diagnosis of the subjects' recordings was performed. Each of the subjects' recordings consisted of multiple 1 min segments. The AHI is commonly used clinically as an indicator of whether a subject is suffering from SA. An individual is considered to have SA if the subject's AHI is >5 (Ruehland et al., 2009). The formula for calculating the AHI is as follows:

$$AHI = \frac{60 \times N}{T} \tag{20}$$

where T is the number of 1 min segments and N indicates the number of 1 min SA segments.

In the per-recording detection, the accuracy, sensitivity, specificity, and AUC of the CNN-BiGRU model were calculated on the retention set as 97.1, 95.7, 100, and 0.996%, respectively. The accuracy was 97.1% because the model misclassified one from the total 35 patients. More specifically, one subject (x12) with SA had an AHI of 6.75, whereas the proposed model calculated an AHI of 4.00, thus classifying the patient as normal. It is worth noting that the low precision per-segment approach may show better per-recording performance because of the relatively small amount of data in the test set (Wang T. et al., 2019). Therefore, according to the previous literature (Sharma

TABLE 3 Per-minute detection performance comparison on Apnea-ECG dataset.

References	Accuracy (%)	Recall (%)	Specificity (%)	Precision (%)	F_1 score (%)
Bahrami and Forouzanfar (2021)	80.7	75.0	84.1	–	74.72
Sharma and Sharma (2016)	83.4	79.5	88.4	–	–
Li et al. (2018)	84.7	88.9	88.4	–	–
Feng et al. (2020)	85.1	86.2	84.4	77.2	81.4
Song et al. (2015)	86.2	82.6	88.4	–	–
Wang T. et al. (2019)	87.6	83.1	90.3	–	–
Sharan et al. (2020)	88.2	82.7	91.6	–	–
Yang et al. (2022)	90.3	87.6	91.9	–	87.3
Chen et al. (2021)	90.6	86.0	93.5	–	87.6
This work (mean)	91.2	86.5	94.2	90.2	88.3

TABLE 4 Per-recording detection comparison on Apnea-ECG dataset.

References	Accuracy (%)	Recall (%)	Specificity (%)	MAE	Corr
Song et al. (2015)	97.1	95.8	100	–	0.860
Sharma and Sharma (2016)	97.1	95.8	100	–	0.841
Li et al. (2018)	100	100	100	9.41	–
Wang T. et al. (2019)	97.1	100	91.7	–	0.943
Feng et al. (2020)	97.1	95.7	100	5.60	–
Shen et al. (2021)	100	100	100	4.23	–
Chen et al. (2021)	100	100	100	–	0.979
Yang et al. (2022)	100	100	100	2.70	0.985
This work	97.1	95.7	100	2.49	0.984

and Sharma, 2016; Wu et al., 2021; Yang et al., 2022), the Pearson correlation coefficient (Corr) and mean absolute error (MAE) were also used as new evaluation indicators to ensure the reliability of the comparison. These metrics are defined as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|AHI_{pre}^{i} - AHI_{\text{true}}^{i}\right| \tag{21}$$

$$Corr = \frac{\sum_{i=1}^{N}\left(AHI_{pre}^{i} - \overline{AHI}_{pre}\right)\left(AHI_{true}^{i} - \overline{AHI}_{true}\right)}{\sqrt{\sum_{i=1}^{N}\left(AHI_{pre}^{i} - \overline{AHI}_{pre}\right)^{2}}\sqrt{\sum_{i=1}^{N}\left(AHI_{true}^{i} - \overline{AHI}_{true}\right)^{2}}} \tag{22}$$

where N is the number of recordings, and AHI_{pre}^{i} and AHI_{true}^{i} represent the predicted and true AHI values of the i-th recording, respectively.

Table 4 lists the comparison of the per-recording classification performance between the CNN-BiGRU model and state-of-the-art works in recent years. As mentioned above, traditional evaluation metrics do not provide a comprehensive and accurate assessment of model performance, and a better approach is to use MAE and Corr metrics. As listed in Table 4, our model achieved 2.49 and 0.984 for the MAE and Corr metrics, respectively. Our model achieved the best performance in terms of MAE metrics. On the Corr metric, literature (Yang et al., 2022) achieved the best value of 0.985, while our model achieved 0.984, which is a comparable result. Overall, our proposed model provides more competitive performance than those of the works presented in Table 4.

3.3. Classification performance on UCDDB dataset

Usually, UCDDB is utilized to evaluate the robustness of the SA detection models (Wang T. et al., 2019; Mashrur et al., 2021; Yang et al., 2022). Similarly, we evaluated our CNN-BiGRU model on UCDDB to demonstrate the model's robustness. Different from the Apnea-ECG dataset, UCDDB was not divided into the training set and test set by the original publishers. As a result, the previous works (Wang T. et al., 2019; Mashrur et al., 2021; Yang et al., 2022) had used their own splitting of training and testing sets in evaluations. In this paper, we used the same preprocessing method for the UCDDB as mentioned in Section 2.2. The difference is that the UCDDB is divided into a training set, a validation set and a test set with a proportion of 8:1:1. Due to the relatively small number of patients with SA at the UCDDB, the training set was balanced by oversampling the minority class (SA events). Meanwhile, per-recording testing is not performed for the same reason.

We used the model trained on the Apnea-ECG dataset to continue training on the UCDDB training set, with the experimental settings mentioned in Section 2.4. On the UCDDB test set, the performance of the CNN-BiGRU model on the accuracy, recall, specificity, and AUC metrics reached 92.3, 70.5, 93.9, and 0.890%, respectively. Figure 5 shows the ROC curves and AUC of the proposed model for per-segment detection.

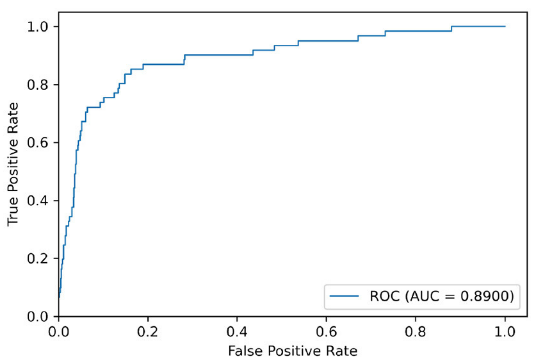


FIGURE 5
ROC curves on UCDDB when positive class is apnea.

Table 5 lists the results of CNN-BiGRU on the test set and compares them with other detection algorithms in the literature. The results show that the CNN-BiGRU model is far superior to the previous models, with an accuracy and specificity of 92.3 and 93.9%, respectively. In regard to recall metrics, we obtained a comparative result to the works (Mashrur et al., 2021). Compared to the Apnea-ECG dataset, our model has a significant decrease in the recall metric on the UCDDB. A major reason for this is that the ratio of pre-processed SA segments to all segments is about 1%, indicating that the class imbalance is intensified. Note that it is a rough comparison in Table 5, as there is no uniform data partitioning of training set and test set for UCDDB. In summary, our CNN-BiGRU model is useful for SA detection.

4. Discussion

4.1. Hyperparameter tuning

In order to verify the efficacy of spatio-temporal blocks, the number of spatio-temporal blocks was tuned from 1 to 5 on PhysioNet Apnea-ECG dataset. As shown in Figure 6 and Table 6, one spatio-temporal block model cannot effectively extract high-level spatio-temporal information. Meanwhile, too many spatio-temporal blocks also fail to learn high-level feature information due to the overfitting problem. Considering that Apnea-ECG dataset suffers from class imbalance, the F_1 score became the main metric we considered. And the CNN-BiGRU model using three spatio-temporal blocks reached the highest values of 88.30% for F_1 score. Therefore, we set the number of spatio-temporal blocks for CNN-BiGRU to 3 in this study.

4.2. Ablation study

We conducted an ablation study on the Apnea-ECG dataset, considering that there was sufficient data in the Apnea-ECG

TABLE 5 Per-minute detection performance comparison on UCDDB dataset.

References	Accuracy (%)	Recall (%)	Specificity (%)	Precision (%)	F_1 score
Wang T. et al. (2019)	71.8	26.6	86.9	–	–
Papini et al. (2018)	74.7	50.6	84.0	–	–
Yang et al. (2022)	75.1	61.1	80.8	–	–
Willemen et al. (2015)	75.9	62.3	–	41.1	–
Xie and Minn (2012)	77.7	69.8	80.3	–	–
Fatimah et al. (2020)	80.4	68.9	87.6	–	–
Mashrur et al. (2021)	81.9	71.6	86.1	–	69.63
This work	92.3	70.5	93.9	46.7	76.0

dataset to fully evaluate the performance of the model. As shown previously, the CNN-BiGRU model uses a convolutional layer, spatio-temporal blocks, and an attention layer to extract features. Therefore, the results of the ablation experiments with the convolutional layer and the attention layer removed separately are listed in Table 7. It was found that removing either the convolutional layer or the attention layer will make the classification performance degrade. Specifically, the accuracy of the models with the convolutional layer removed and the attention layer removed is decreased by 0.47 and 0.67%, respectively. Overall, using the convolutional layer and attention layer improved the classification performance of the CNN-BiGRU model.

4.3. Cross-dataset evaluation

Cross-dataset evaluations are performed to demonstrate the general performance of our proposed model, using the Apnea-ECG and UCDDB datasets. Specifically, the model is trained on one dataset and evaluated directly on another dataset. When CNN-BiGRU was trained on Apnea-ECG and tested on UCDDB, an accuracy of 85.3% and an F1-score of 50.5% were achieved. Similarly, when it was trained on UCDDB and evaluated on Apnea-ECG, the accuracy was 53.8% and the F1-score was 36.3%. It is found that the performances of cross-dataset evaluation are not satisfactory. To comprehensively understand the evaluation, we performed the same cross-dataset evaluation using a previous state-of-the-art model (Chen et al., 2021) listed in Table 3. The accuracy achieved was 85.9% and the F1-score was 51.1% using the UCDDB as the testing set and Apnea-ECG as the training set. They are very slightly better than those of our model (85.9 vs. 85.3%; 51.1 vs. 50.5%). However, when it was trained with UCDDB and tested on Apnea-ECG, the accuracy and the F1-score were 45.5 and 29.2%, respectively. Obviously, our model has outperformed this previous model (53.8 vs. 45.5%; 36.3 vs. 29.2%). In general, CNN-BiGRU is superior to the compared model (Chen et al., 2021), in terms of cross-dataset evaluation.

Finally, we attribute the low performance of cross-dataset evaluation to the following reasons: (1) the populations of datasets are different. For example, subjects with central apnea and periodic respiratory episodes are included in UCDDB; (2) the different sampling rates may impact the performance (the ECG signals are sampled at 100 Hz on Apnea-ECG while 128 Hz on UCDDB); (3) UCDDB has a severe class imbalance problem. In other words, the distributions of normal and SA are quite different between the two datasets.

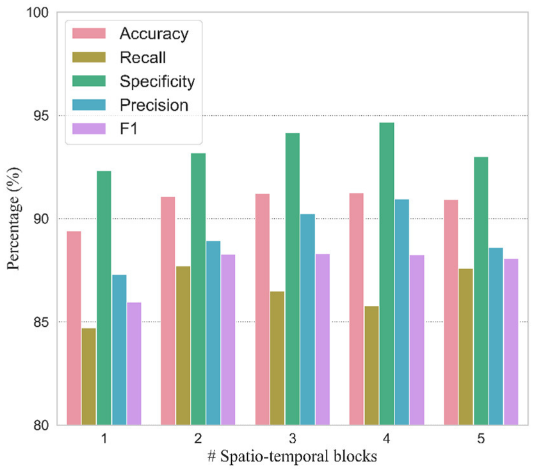


FIGURE 6
Hyperparameter tuning for the number of spatio-temporal blocks.

5. Conclusion

In this study, a novel spatio-temporal learning-based model named CNN-BiGRU was explored to classify SA events from ECG signals. Specifically, the proposed CNN-BiGRU is an end-to-end deep learning model, which consists of multiple spatio-temporal blocks. Each block has the identical architecture with a CNN layer, a max-pooling layer, and a BiGRU layer. This architecture with repeated spatio-temporal blocks can well capture the morphological spatial feature information as well as the temporal feature information from ECG signals.

TABLE 6 Comparison of per-segment detection results using different numbers of spatio-temporal blocks.

# Spatio-temporal blocks	Accuracy (%)	Recall (%)	Specificity (%)	Precision (%)	F_1 score
1	89.40 ± 0.4486	84.69 ± 1.867	92.32 ± 1.135	87.30 ± 1.509	85.95 ± 0.6510
2	91.08 ± 0.2314	87.70 ± 1.449	93.19 ± 0.9452	88.92 ± 1.262	88.28 ± 0.3309
3	91.22 ± 0.2098	86.48 ± 1.360	94.16 ± 0.9404	90.23 ± 1.316	88.30 ± 0.2833
4	91.26 ± 0.1953	85.77 ± 1.588	94.67 ± 0.8665	90.94 ± 1.243	88.26 ± 0.3714
5	90.92 ± 0.1502	87.59 ± 1.222	92.99 ± 0.7195	88.60 ± 0.9142	88.08 ± 0.2603

TABLE 7 Ablation of CNN-BiGRU.

Convolutional layer	Attention layer	Accuracy (%)	Recall (%)	Specificity (%)	Precision (%)	F_1 score
✓		90.55 ± 0.1128	85.28 ± 1.741	93.81 ± 1.183	89.60 ± 1.629	87.35 ± 0.2002
	✓	90.75 ± 0.2061	86.57 ± 2.054	93.35 ± 1.118	89.05 ± 1.468	87.76 ± 0.4489
✓	✓	91.22 ± 0.2098	86.48 ± 1.360	94.16 ± 0.9404	90.23 ± 1.316	88.30 ± 0.2833

Experiment results on the apnea-ECG dataset showed that the proposed CNN-BiGRU achieved an accuracy of 91.22 and 97.10% for per-minute classification and per-recording classification, respectively. And the accuracy on the UCDDB dataset reached 91.24%. In contrast to the previous state-of-the-art methods, our proposed CNN-BiGRU has an obvious advantage with a big margin. It means that the CNN-BiGRU can be potentially deployed into a medical system to help physicians to screen out SA patients to avoid malignant events. In future work, we will further apply the proposed model to real healthcare systems.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: PhysioNet Apnea-ECG: https://physionet.org/content/apnea-ecg/1.0.0/; UCDDB: https://archive.physionet.org/physiobank/database/ucddb/.

Author contributions

Conceptualization: WZ and WM. Methodology: JC and MS. Software: JC. Writing: WZ and JC. Project administration: WZ. All authors contributed to the article and approved the submitted version.

Funding

This work was partly supported by the National Key R&D Program of China under Grant 2019YFB1804003.

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

References

Ali, S. Q., Khalid, S., and Brahim Belhaouari, S. (2019). A novel technique to diagnose sleep apnea in suspected patients using their ECG data. *IEEE Access* 7, 35184–35194. doi: 10.1109/ACCESS.2019.2904601

Ankhili, A., Tao, X., Cochrane, C., Coulon, D., and Koncar, V. (2018). Washable and reliable textile electrodes embedded into underwear fabric for electrocardiography (ECG) monitoring. *Materials* 11:256. doi: 10.3390/ma11020256

Bahrami, M., and Forouzanfar, M. (2021). "Detection of sleep apnea from single-lead ECG: comparison of deep learning algorithms," in *2021 IEEE International Symposium on Medical Measurements and Applications* (Lausanne: IEE), 1–5. doi: 10.1109/MeMeA52024.2021.9478745

Baty, F., Boesch, M., Widmer, S., Annaheim, S., Fontana, P., Camenzind, M., et al. (2020). Classification of sleep apnea severity by electrocardiogram monitoring using a novel wearable device. *Sensors* 20:286. doi: 10.3390/s20010286

Benjafield, A. V., Ayas, N. T., Eastwood, P. R., Heinzer, R., Ip, M. S., Morrell, M. J., et al. (2019). Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *Lancet Respir. Med.* 7, 687–698. doi: 10.1016/S2213-2600(19)30198-5

Bsoul, M., Minn, H., and Tamil, L. (2011). Apnea medassist: real-time sleep apnea monitor using single-lead ECG. *IEEE Trans. Inform. Technol. Biomed.* 15, 416–427. doi: 10.1109/TITB.2010.2087386

Chen, L., Zhang, X., and Song, C. (2015). An automatic screening approach for obstructive sleep apnea diagnosis based on single-lead electrocardiogram. *IEEE Trans. Autom. Sci. Eng.* 12, 106–115. doi: 10.1109/TASE.2014.2345667

Chen, X., Chen, Y., Ma, W., Fan, X., and Li, Y. (2021). "SE-MSCNN: a lightweight multi-scaled fusion network for sleep apnea detection using single-lead ECG signals," in *2021 IEEE International Conference on Bioinformatics and Biomedicine* (Houston, TX: IEEE), 1276–1280. doi: 10.1109/BIBM52615.2021.9669358

Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., et al. (2014). "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (Doha: Association for Computational Linguistics), 1724–1734. doi: 10.3115/v1/D14-1179

Chung, J., Çaglar Gülçehre, Cho, K., and Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *ArXiv: abs/1412.3555*. doi: 10.48550/arXiv.1412.3555

Fan, X., Wang, H., Xu, F., Zhao, Y., and Tsui, K.-L. (2019). Homecare-oriented intelligent long-term monitoring of blood pressure using electrocardiogram signals. *IEEE Trans. Indus. Inform.* 16, 7150–7158. doi: 10.1109/TII.2019.2962546

Fan, X., Yao, Q., Cai, Y., Miao, F., Sun, F., and Li, Y. (2018). Multiscaled fusion of deep convolutional neural networks for screening atrial fibrillation from single lead short ECG recordings. *IEEE J. Biomed. Health Inform.* 22, 1744–1753. doi: 10.1109/JBHI.2018.2858789

Fatimah, B., Singh, P., Singhal, A., and Pachori, R. B. (2020). Detection of apnea events from ECG segments using Fourier decomposition method. *Biomed. Signal Process. Control* 61:102005. doi: 10.1016/j.bspc.2020.102005

Faust, O., Acharya, U. R., Ng, E., and Fujita, H. (2016). A review of ECG-based diagnosis support systems for obstructive sleep apnea. *J. Mech. Med. Biol.* 16:1640004. doi: 10.1142/S0219519416400042

Feng, K., Qin, H., Wu, S., Pan, W., and Liu, G. (2020). A sleep apnea detection method based on unsupervised feature learning and single-lead electrocardiogram. *IEEE Trans. Instrumen. Measure.* 70, 1–12. doi: 10.1109/TIM.2020.3017246

Gaiduk, M., Orcioni, S., Conti, M., Seepold, R., Penzel, T., Madrid, N. M., et al. (2020). "Embedded system for non-obtrusive sleep apnea detection*," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society* (Montreal, QC: IEEE), 2776–2779. doi: 10.1109/EMBC44109.2020.9176075

Graco, M., Schembri, R., Cross, S., Thiyagarajan, C., Shafazand, S., Ayas, N. T., et al. (2018). Diagnostic accuracy of a two-stage model for detecting obstructive sleep apnoea in chronic tetraplegia. *Thorax* 73, 864–871. doi: 10.1136/thoraxjnl-2017-211131

Gutiérrez-Tobal, G. C., Álvarez, D., Gomez-Pilar, J., Campo, F. d., and Hornero, R. (2015). Assessment of time and frequency domain entropies to detect sleep apnoea in heart rate variability recordings from men and women. *Entropy* 17, 123–141. doi: 10.3390/e17010123

Hamilton, P. (2002). "Open source ECG analysis," in *Computers in Cardiology*, (Memphis, TN: IEEE), 101–104. doi: 10.1109/CIC.2002.1166717

John, A., Cardiff, B., and John, D. (2021). "A 1D-CNN based deep learning technique for sleep apnea detection in IoT sensors," in *2021 IEEE International Symposium on Circuits and Systems* (Daegu: IEEE), 1–5. doi: 10.1109/ISCAS51556.2021.9401300

King, S., and Cuellar, N. (2016). Obstructive sleep apnea as an independent stroke risk factor: a review of the evidence, stroke prevention guidelines, and implications for neuroscience nursing practice. *J. Neurosci. Nurs.* 48, 133–142. doi: 10.1097/JNN.0000000000000196

Kingma, D. P., and Ba, J. (2017). Adam: a method for stochastic optimization. *arXiv [Preprint]*. arXiv: 1412.6980. doi: 10.48550/arXiv.1412.6980

Li, K., Pan, W., Li, Y., Jiang, Q., and Liu, G. (2018). A method to detect sleep apnea based on deep neural network and hidden Markov model using single-lead ECG signal. *Neurocomputing* 294, 94–101. doi: 10.1016/j.neucom.2018.03.011

Lin, C.-H., Lurie, R. C., and Lyons, O. D. (2020). Sleep apnea and chronic kidney disease: a state-of-the-art review. *Chest* 157, 673–685. doi: 10.1016/j.chest.2019.09.004

Luong, T., Pham, H., and Manning, C. D. (2015). "Effective approaches to attention-based neural machine translation," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (Lisbon: Association for Computational Linguistics), 1412–1421. doi: 10.18653/v1/D15-1166

Lyons, O. D., and Ryan, C. M. (2015). Sleep apnea and stroke. *Can. J. Cardiol.* 31, 918–927. doi: 10.1016/j.cjca.2015.03.014

Mashrur, F. R., Islam, M. S., Saha, D. K., Islam, S. R., and Moni, M. A. (2021). SCNN: Scalogram-based convolutional neural network to detect obstructive sleep apnea using single-lead electrocardiogram signals. *Comput. Biol. Med.* 134:104532. doi: 10.1016/j.compbiomed.2021.104532

Mostafa, S. S., Mendonça, F., Morgado-Dias, F., and Ravelo-García, A. (2017). "Spo2 based sleep apnea detection using deep learning," in *2017 IEEE 21st International Conference on Intelligent Engineering Systems* (Larnaca: IEEE), 91–96. doi: 10.1109/INES.2017.8118534

Mostafa, S. S., Morgado-Dias, F., and Ravelo-García, A. G. (2020). Comparison of SFS and MRMR for oximetry feature selection in obstructive sleep apnea detection. *Neural Comput. Appl.* 32, 15711–15731. doi: 10.1007/s00521-018-3455-8

Papini, G. B., Fonseca, P., Margarito, J., van Gilst, M. M., Overeem, S., Bergmans, J. W., et al. (2018). "On the generalizability of ECG-based obstructive sleep apnea monitoring: merits and limitations of the apnea-ECG database," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Honolulu, HI: IEEE), 6022–6025. doi: 10.1109/EMBC.2018.8513660

Penzel, T., Moody, G. B., Mark, R. G., Goldberger, A. L., and Peter, J. H. (2000). "The apnea-ECG database," in *Computers in Cardiology 2000. Vol. 27* (Cambridge, MA: IEEE), 255–258. doi: 10.1109/CIC.2000.898505

Rosen, I. M., Kirsch, D. B., Chervin, R. D., Carden, K. A., Ramar, K., Aurora, R. N., et al. (2017). Clinical use of a home sleep apnea test: an American academy of sleep medicine position statement. *J. Clin. Sleep Med.* 13, 1205–1207. doi: 10.5664/jcsm.6774

Ruehland, W. R., Rochford, P. D., O'Donoghue, F. J., Pierce, R. J., Singh, P., and Thornton, A. T. (2009). The new AASM criteria for scoring hypopneas: impact on the apnea hypopnea index. *Sleep* 32, 150–157. doi: 10.1093/sleep/32.2.150

Russell, M. B., Kristiansen, H. A., and Kværner, K. J. (2014). Headache in sleep apnea syndrome: epidemiology and pathophysiology. *Cephalalgia* 34, 752–755. doi: 10.1177/0333102414538551

Sharan, R. V., Berkovsky, S., Xiong, H., and Coiera, E. (2020). "ECG-derived heart rate variability interpolation and 1-d convolutional neural networks for detecting sleep apnea," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine* & *Biology Society* (Montreal, QC: IEEE), 637–640. doi: 10.1109/EMBC44109.2020.9175998

Sharma, H., and Sharma, K. (2016). An algorithm for sleep apnea detection from single-lead ECG using hermite basis functions. *Comput. Biol. Med.* 77, 116–124. doi: 10.1016/j.compbiomed.2016.08.012

Shen, Q., Qin, H., Wei, K., and Liu, G. (2021). Multiscale deep neural network for obstructive sleep apnea detection using RR interval from single-lead ECG signal. *IEEE Trans. Instrumen. Measure.* 70, 1–13. doi: 10.1109/TIM.2021.3062414

Smruthy, A., and Suchetha, M. (2017). Real-time classification of healthy and apnea subjects using ECG signals with variational mode decomposition. *IEEE Sensors J.* 17, 3092–3099. doi: 10.1109/JSEN.2017.2690805

Somers, V. K., White, D. P., Amin, R., Abraham, W. T., Costa, F., Culebras, A., et al. (2008). Sleep apnea and cardiovascular disease: an American heart association/American college of cardiology foundation scientific statement from the American heart association council for high blood pressure research professional education committee, council on clinical cardiology, stroke council, and council on cardiovascular nursing in collaboration with the national heart, lung, and blood institute national center on sleep disorders research (national institutes of health). *J. Am. Coll. Cardiol.* 52, 686–717. doi: 10.1016/j.jacc.2008.05.002

Song, C., Liu, K., Zhang, X., Chen, L., and Xian, X. (2015). An obstructive sleep apnea detection approach using a discriminative hidden Markov model from ECG signals. *IEEE Trans. Biomed. Eng.* 63, 1532–1542. doi: 10.1109/TBME.2015.2498199

Vanek, J., Prasko, J., Genzor, S., Ociskova, M., Kantor, K., Holubova, M., et al. (2020). Obstructive sleep apnea, depression and cognitive impairment. *Sleep Med.* 72, 50–58. doi: 10.1016/j.sleep.2020.03.017

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). "Attention is all you need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17* (Red Hook, NY: Curran Associates Inc.), 6000–6010.

Viswabhargav, C. S., Tripathy, R., and Acharya, U. R. (2019). Automated detection of sleep apnea using sparse residual entropy features with various dictionaries extracted from heart rate and EDR signals. *Comput. Biol. Med.* 108, 20–30. doi: 10.1016/j.compbiomed.2019.03.016

Wang, L., Lin, Y., and Wang, J. (2019). A RR interval based automated apnea detection approach using residual network. *Comput. Methods Prog. Biomed.* 176, 93–104. doi: 10.1016/j.cmpb.2019.05.002

Wang, T., Lu, C., Shen, G., and Hong, F. (2019). Sleep apnea detection from a single-lead ECG signal with automatic feature-extraction through a modified lenet-5 convolutional neural network. *PeerJ* 7:e7731. doi: 10.7717/peerj.7731

Wang, Z., Wang, X., and Wang, G. (2018). Learning fine-grained features *via* a CNN tree for large-scale classification. *Neurocomputing* 275, 1231–1240. doi: 10.1016/j.neucom.2017.09.061

Weder, M., Hegemann, D., Amberg, M., Hess, M., Boesel, L. F., Abächerli, R., et al. (2015). Embroidered electrode with silver/titanium coating for long-term ECG monitoring. *Sensors* 15, 1750–1759. doi: 10.3390/s150101750

Willemen, T., Varon, C., Dorado, A. C., Haex, B., Vander Sloten, J., and Van Huffel, S. (2015). Probabilistic cardiac and respiratory based classification of sleep and apneic events in subjects with sleep apnea. *Physiol. Measure*. 36:2103. doi: 10.1088/0967-3334/36/10/2103

Wu, Y., Pang, X., Zhao, G., Yue, H., Lei, W., and Wang, Y. (2021). A novel approach to diagnose sleep apnea using enhanced frequency extraction network. *Comput. Methods Prog. Biomed*. 206:106119. doi: 10.1016/j.cmpb.2021.106119

Xie, B., and Minn, H. (2012). Real-time sleep apnea detection by classifier combination. *IEEE Trans. Inform. Technol. Biomed*. 16, 469–477. doi: 10.1109/TITB.2012.2188299

Yang, Q., Zou, L., Wei, K., and Liu, G. (2022). Obstructive sleep apnea detection from single-lead electrocardiogram signals using one-dimensional squeeze-and-excitation residual group network. *Comput. Biol. Med*. 140:105124. doi: 10.1016/j.compbiomed.2021.105124

Zhang, Y., Yao, S., Yang, R., Liu, X., Qiu, W., Han, L., et al. (2022). Epileptic seizure detection based on bidirectional gated recurrent unit network. *IEEE Trans. Neural Syst. Rehabil. Eng*. 30, 135–145. doi: 10.1109/TNSRE.2022.3143540

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 10 August 2022
DOI 10.3389/fnins.2022.974712

OPEN ACCESS

EDITED BY
Weidong Gao,
Beijing University of Posts
and Telecommunications (BUPT),
China

REVIEWED BY
Shixiong Chen,
Shenzhen Institutes of Advanced
Technology (CAS), China
Bolin Liu,
Xi'an International Medical Center
Hospital, China

*CORRESPONDENCE
Weiqun Song
songwq66@126.com

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

RECEIVED 21 June 2022
ACCEPTED 22 July 2022
PUBLISHED 10 August 2022

CITATION
Zhang Y, Ye L, Cao L and Song W
(2022) Resting-state
electroencephalography changes
in poststroke patients with visuospatial
neglect.
Front. Neurosci. 16:974712.
doi: 10.3389/fnins.2022.974712

Resting-state electroencephalography changes in poststroke patients with visuospatial neglect

Yichen Zhang, Linlin Ye, Lei Cao and Weiqun Song*

Department of Rehabilitation, Xuanwu Hospital, Capital Medical University, Beijing, China

Background: This study aimed to explore the electrophysiological characteristics of resting-state electroencephalography (rsEEG) in patients with visuospatial neglect (VSN) after stroke.

Methods: A total of 44 first-event sub-acute strokes after right hemisphere damage (26 with VSN and 18 without VSN) were included. Besides, 18 age-matched healthy participants were used as healthy controls. The resting-state electroencephalography (EEG) of 64 electrodes was recorded to obtain the power of the spectral density of different frequency bands. The global delta/alpha ratio (DAR), DAR over the affected hemispheres (DAR_{AH}), DAR over the unaffected hemispheres (DAR_{UH}), and the pairwise-derived brain symmetry index (pdBSI; global and four bands) were compared between groups and receiver operating characteristic (ROC) curve analysis was conducted. The Barthel index (BI), Fugl-Meyer motor function assessment (FMA), and Berg balance scale (BBS) were used to assess the functional state of patients. Visuospatial neglect was assessed using a battery of standardized tests.

Results: We found that patients with VSN performed poorly compared with those without VSN. Analysis of rsEEG revealed increased delta and theta power and decreased alpha and beta power in stroke patients with VSN. Compared to healthy controls and poststroke non-VSN patients, patients with VSN showed a higher DAR ($P < 0.001$), which was significantly positively correlated with the BBS (DAR: $r = -0.522$, $P = 0.006$; DAR_{AH}: $r = -0.521$, $P = 0.006$; DAR_{UH}: $r = -0.494$, $P = 0.01$). The line bisection task was positively correlated with DAR ($r = 0.458$, $P = 0.019$) and DAR_{AH} ($r = 0.483$, $P = 0.012$), while the star cancellation task was only positively correlated with DAR_{AH} ($r = 0.428$, $P = 0.029$). DAR_{AH} had the best discriminating value between VSN and non-VSN, with an area under the curve (AUC) of 0.865. Patients with VSN showed decreased alpha power in the parietal and occipital areas of the right hemisphere. A higher parieto-occipital pdBSI$_{alpha}$ was associated with a worse line bisection task (r = 0.442, P = 0.024).

Conclusion: rsEEG may be a useful tool for screening for stroke patients with visuospatial neglect, and DAR and parieto-occipital pdBSI$_{alpha}$ may be useful biomarkers for visuospatial neglect after stroke.

KEYWORDS

stroke, visuospatial neglect, resting-state EEG, delta/alpha ratio, alpha oscillation

Introduction

Visuospatial neglect (VSN) is the most frequent neglect syndrome, characterized by failure to orient or respond to visual stimuli presented in the contralesional space (Heilman and Valenstein, 1979; Parton et al., 2004), particularly in patients with right hemisphere damage (Ten Brink et al., 2017). The prevalence of VSN after unilateral stroke is 30% (Esposito et al., 2020), and it disrupts basic activities of daily living (such as dressing and walking; Nijboer et al., 2013; Aravind and Lamontagne, 2018) and increases the risk of falls (Chen et al., 2015). Worsely, many individuals with VSN are unaware of these deficits. Furthermore, VSN hinders the ability to perform rehabilitation and limits recovery during early post-stroke neuroplasticity enhancement. VSN is an important predictor of poor functional recovery 1 year after stroke (Jehkonen et al., 2000; Hammerbeck et al., 2019). Despite its high prevalence and serious sequelae, VSN often remains unrecognized and undertreated because of the heterogeneity of its clinical manifestations and the limitations of assessment methods (Puig-Pijoan et al., 2018). Pencil-and-paper clinical tests are the most commonly used tools to assess VSN, but they sometimes lack sensitivity (Azouvi, 2017). Some patients with severe VSN symptoms may be unable to complete the scale assessment, whereas patients with mild VSN symptoms may have normal scale assessments. Therefore, in addition to assessing the severity of VSN symptoms from a behavioral perspective, new assessment tools that can sensitively reflect dynamic changes at physiological levels are needed.

Electroencephalography (EEG) is a non-invasive method with high temporal resolution, which contributes to the rapid evaluation of instantaneous brain function. Closed-eye resting-state EEG (rsEEG) provides an important opportunity to examine EEG oscillatory patterns of spontaneous brain activity unbiased by any task (Fingelkurts and Fingelkurts, 2015). Specific EEG oscillation patterns are associated with specific psychological or behavioral states (Herrmann et al., 2016). A previous study showed that alpha desynchronization in the spatially contralateral hemisphere of attention is a reliable marker of attentional orientation in the healthy human brain (Lasaponara et al., 2019). A large number of studies have found that spectral rsEEG can be a useful tool for auxiliary diagnosis of Alzheimer's disease, post-stroke aphasia, and post-stroke depression. However, few studies have used rsEEG as a VSN screening tool. Previous studies have found that resting EEG topography has high sensitivity and reliability, and can help distinguish patients with different severities of VSN (Pirondini et al., 2020). However, there are no studies on the spectral power parameters of rsEEG.

The delta/alpha ratio (DAR) and pairwise-derived brain symmetry index (pdBSI) are commonly used resting-state EEG parameters, which are potentially valuable early predictors of the severity of post-stroke dysfunction (Saes et al., 2020, 2021). Hemispheric stroke has been associated with increased low-frequency oscillations in delta bands (Cassidy et al., 2020) and decreased alpha activity (Zappasodi et al., 2019). Compared to the individual spectral components, the DAR quantifying these spectral power changes may more sensitively reflect the severity of neurological deficits. The pdBSI assesses the asymmetry of the spectral power distribution between hemispheres after unilateral hemispheric stroke by calculating the power spectral densities along with homologous EEG channel pairs (Sheorajpanday et al., 2009). However, it is still unknown whether these rsEEG parameters differ in post-stroke VSN patients compared to healthy subjects and patients without VSN after stroke, and whether they are related to the severity of VSN.

In this study, we aimed to determine the specific rsEEG characteristics in poststroke patients with visuospatial neglect. The rsEEG of patients with VSN after stroke, patients without VSN after stroke and healthy individuals were collected, and the differences in rsEEG parameters (DAR and pdBSI) and topographic maps among the three groups and their correlation with clinical manifestations were analyzed. We aimed to determine whether specific rsEEG features of these post-stroke VSN patients could be used to aid in diagnosis and evaluation and to help design clinical screening procedures for visuospatial neglect in post-stroke patients.

Materials and methods

Participants

Patients were recruited from sequential admissions to the Department of Rehabilitation at the Xuanwu Hospital of Capital Medical University, China. A total of 44 first-event subacute stroke patients after right hemisphere damage were included, comprising of 26 patients with left VSN and 18 patients without VSN (non-VSN). The demographic characteristics are reported in **Tables 1**, **2**. We also recruited 18 age-matched healthy controls (HC). The inclusion criteria for stroke patients were as follows: (1) first-ever cerebral stroke according to computed tomography or magnetic resonance imaging (MRI) scan; (2) ability to complete the necessary checks; (3) age $\geq$ 18 years; and (4) all right-handed patients who had normal or corrected-to-normal visual acuity. The exclusion criteria were as follows: (1) other neurological diseases; (2) severe cognitive problems, that is, Mini-Mental State Examination score $<$ 18; and (3) worsening condition. All participants provided written informed consent and the study was approved by the Ethics Committee of Xuanwu Hospital (approval number: [2020]155).

All stroke patients were assessed for activities of daily living (ADL), motor function, and balance. The Barthel index (BI) was used to assess patients' daily living abilities. The total BI score was 100, and higher scores suggest stronger daily living ability. The Fugl-Meyer Motor Assessment (FMA) Scale included upper and lower extremity movements, with 33 assessment items for upper extremity movement and 17 assessment items for lower extremity movement, with a total score of 100. Higher scores indicated better limb motor function. The Berg Balance Scale (BBS) was a commonly used balance scale that can comprehensively evaluate the dynamic and static balance function of stroke patients in the sitting and standing positions. It consisted of 14 items with a total score of 56. A lower score indicates poorer balance function.

Clinical assessment of visuospatial neglect

Line bisection task

The patients were asked to bisect five horizontal black lines of differing lengths (80, 100, 120, 140, and 160 mm). The deviation of the patient's marked point from the true midpoint of the line (in millimeters) was measured and converted to a percentage score (line bisection error [LBE]). Rightward deviations from the true line center were scored as positive and leftward deviations were scored as negative. VSN was diagnosed when the average LBE was $>$ 12%.

Line cancellation task

This test involved 30 lines, each with a length of 15 mm, evenly distributed on the paper. Stroke patients were asked to cross all the lines on the page. The ratio of the number of missing line segments to 30 was the omission rate for the line cancellation task (ORL).

Clock copying task

The subjects were instructed to copy a clock on paper. Errors of omission in hands and numbers were considered pathological VSN.

Star cancellation task

There were 56 small stars interspersed with 52 large stars, 10 short words, and 13 letters. All patients were asked to mark all small stars. The omission rate of stars (ORS) was the number of missing small stars divided by 56.

Electroencephalography

All subjects were requested to relax and not engage in any specific mental activity during EEG recording. Eye-closed rsEEG signals were recorded for 5 min using a NeuroScan NuAmps amplifier (Compumedics United States, Ltd., El Paso, TX, United States), and 64 Ag–AgCl electrodes were mounted on a Quik Cap using a modified 10–20 placement scheme to record the EEG. The EEG data were recorded with a 0.1–100 Hz band-pass filter at a sampling rate of 1,000 Hz. The ground electrode was placed on the forehead and the reference electrode was placed on the nose. The impedance of all the electrodes was maintained at $\leq 10\ k\Omega$.

Pre-processing

Offline EEG preprocessing was conducted using the open-source EEGLAB toolbox and custom MATLAB 2013b (Math Works, Natick, NA). The raw EEG data were filtered using an FIR filter at 0.1–40 Hz. A 48–52 Hz notch filter was used to eliminate the power frequency interference. The data were then segmented into 2 s epochs. Bad channels were discarded by visual inspection and interpolated using the spherical method, followed by re-referencing to the remaining average. Data portions contaminated by eye blinks and eye movements were corrected using independent component analysis (ICA). The EEG epochs with amplitude values exceeding $\pm$ 100 μV at any electrode were excluded. The

TABLE 1 Stroke characteristics.

Subject number	Gender	Age (years)	Duration after stroke (days)	Type of stroke	Lesion location
VSN01	Female	47	16	Ischemic	T, P, BG
VSN02	Male	48	31	Ischemic	BG
VSN03	Male	62	9	Ischemic	F, BG
VSN04	Male	61	16	Ischemic	F, T, P, BG, CS
VSN05	Male	67	21	Ischemic	F, T, BG
VSN06	Female	69	18	Ischemic	F, T, P, BG, CR
VSN07	Male	69	25	Ischemic	F, T, P, CS
VSN08	Female	66	12	Ischemic	F, T
VSN09	Female	51	18	Ischemic	F, T, P, BG
VSN10	Male	58	36	Ischemic	F, T, P
VSN11	Female	73	19	Ischemic	F, T, P
VSN12	Male	48	17	Hemorrhagic	BG
VSN13	Male	37	42	Hemorrhagic	F, T, P
VSN14	Male	58	17	Ischemic	BG
VSN15	Male	65	13	Ischemic	F, BG
VSN16	Male	61	30	Hemorrhagic	BG
VSN17	Female	31	24	Ischemic	F, T, P, BG
VSN18	Female	71	10	Ischemic	F, BG
VSN19	Female	78	23	Ischemic	F, T, P, BG
VSN20	Male	49	22	Ischemic	F, T, P, BG, CR
VSN21	Male	59	14	Ischemic	F, T, P, BG
VSN22	Female	66	24	Ischemic	BG, CS
VSN23	Male	67	28	Ischemic	F, P, CR
VSN24	Male	73	8	Ischemic	T, BG
VSN25	Male	56	40	Ischemic	F, T, BG
VSN26	Male	67	17	Hemorrhagic	P
Non-VSN01	Male	74	21	Ischemic	BG
Non-VSN02	Male	45	14	Hemorrhagic	F, BG
Non-VSN03	Male	53	15	Ischemic	F, T, P, BG, CR
Non-VSN04	Male	38	19	Ischemic	BG
Non-VSN05	Male	69	14	Ischemic	F, P, BG
Non-VSN06	Male	53	8	Ischemic	BG
Non-VSN07	Male	43	26	Hemorrhagic	F, T, P, BG
Non-VSN08	Female	63	41	Ischemic	BG, CR
Non-VSN09	Male	65	20	Ischemic	F, T, P, CR
Non-VSN10	Female	69	22	Ischemic	F, T, P, BG
Non-VSN11	Male	58	27	Ischemic	BG
Non-VSN12	Male	72	20	Ischemic	BG, CR
Non-VSN13	Male	63	32	Hemorrhagic	F, P, BG
Non-VSN14	Female	66	10	Ischemic	BG
Non-VSN15	Female	68	15	Ischemic	T, P, BG, CS
Non-VSN16	Male	68	21	Ischemic	F, P, BG
Non-VSN17	Male	62	20	Ischemic	F, P
Non-VSN18	Male	57	18	Hemorrhagic	F, T, P, BG

BG, Basal ganglia; CR, Corona radiata; CS, Centrum semiovale; F, Frontal lobe; T, Temporal lobe; P, Parietal lobe.

power of spectral density (PSD, μV^2/Hz) using fast Fourier transform (FFT) was carried out for four frequency bands: delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), and beta (13–30 Hz).

Resting-state electroencephalography parameters

Delta/alpha ratio

The DAR was defined as the ratio between the mean delta power and mean alpha power. For every channel *c*, the power of the delta and alpha (f = 1,. . .,4 Hz and f = 8,. . .,12 Hz, respectively) was determined as the mean of the spectral power [$P_c(f)$]over this range. The DAR was computed as

$$DAR_c = \frac{\langle P_c\left(f\right)\rangle_{f=1,\ldots,4Hz}}{\langle P_c\left(f\right)\rangle_{f=8,\ldots,12Hz}}$$

The ratios were averaged over all N EEG channels yielding the global DAR as:

$$DAR = \frac{1}{N}\sum_{c=1}^{N} DAR_c$$

TABLE 2 Patients' demographic and clinical characteristics.

	VSN (n = 26)	Non-VSN (n = 18)	HC (n = 18)	p
Gender(female/male)	9/17	4/14	5/13	0.666
Age($M \pm$ SD)	59.88±11.40	60.33±10.35	54.72±9.10	0.197
Type of stroke (ischemic/hemorrhagic)	22/4	14/4	–	0.857
Lesion location (cortical/cortico-subcortical/subcortical)	1/21/4	1/10/7	–	0.185
Poststroke time($M\pm$ SD; Day)	21.15±9.04	20.17±7.85	–	0.709
BI	34.23±15.28	53.06±19.41	–	0.001
FMA	30.27±18.84	50.11±16.47	–	0.004
BBS	8.19±8.20	16.39±13.55	–	0.016

In addition, the DAR was calculated over the affected hemisphere (DAR_{AH}) and the unaffected hemisphere (DAR_{UH}), excluding the electrodes covering the midline.

Pairwise-derived brain symmetry index

The pdBSI was defined as the absolute pairwise normalized difference in spectral power between the homologous channels C_L and C_R for the left and right hemispheres, excluding the electrodes covering the midline. The difference was averaged over a range of 1–25 Hz, according to

$$BSI_{CP} = \langle \left| \frac{P_{C_R}(f) - P_{C_L}(f)}{P_{C_R}(f) + P_{C_L}(f)} \right| \rangle f = 1, \ldots, 25Hz$$

These values were averaged over all channel pairs (cp):

$$BSI = \frac{2}{N} \sum_{CP=1}^{N/2} BSI_{CP}$$

The pdBSI estimated the global asymmetry along with homologous channel pairs, ranging from 0 to 1, with 0 defined as the maximal symmetry. In addition, the BSI values of each frequency band and of the alpha frequency band of the parieto-occipital region (Pz, P1, P2, P3, P4, P5, P6, P7, P8, POz, PO3, PO4, PO5, PO6, PO7, PO8, Oz, O1, O2) were also calculated.

Statistical analysis

All statistical analyses were performed using IBM SPSS Statistics V22.0 (IBM Corp, Armonk, NY, United States). The distribution of data was tested for normality using the Kolmogorov–Smirnov test. Categorical data were analyzed using the chi-squared test. One-way ANOVA with *post hoc* testing using Bonferroni's test was used to test the differences among the three groups (VSN, non-VSN, and HC). Differences between the VSN and non-VSN groups were compared using an independent sample t-test. Pearson correlation analysis was performed to investigate the relationship between rsEEG parameters and patient characteristics. In addition, EEG parameters that showed a significant difference among groups were analyzed using receiver operating characteristic (ROC) curve analysis. The sensitivity/specificity cut-off values, positive predictive value (PPV), and negative predictive value (NPV) were determined using Youden's index.

Results

Demographic information and descriptive data

The relevant demographic and clinical characteristics of the three groups were shown in **Table 2**. There were no significant differences in age or sex among the three groups. No significant difference was found in the type of stroke or time since stroke onset between VSN and non-VSN subjects. The BI, FMA, and BBS scores of the VSN patients were significantly lower than those of the non-VSN patients.

According to the Pearson correlation analysis results, LBE scores were significantly correlated with BBS scores (r = –0.605, P = 0.001) but not with BI or FMA scores in patients with VSN. Scatter plots showing the relationship between LBE and BBS are presented in **Figure 1**.

Electroencephalography parameters

Delta/alpha ratio

The ANOVA results shown in **Figure 2** indicated that there were significant differences in the DAR among the three groups. DAR values were higher in patients with VSN compared with both patients without VSN and HCs (F = 28.348, P < 0.001; VSN vs HC: P < 0.001; VSN vs Non-VSN: P < 0.001). Patients with non-VSN had a higher DAR than HCs (P = 0.048). Patients with VSN had higher DAR_{AH} and DAR_{UH} values than non-VSN patients.

For the DAR, a trend toward a negative association with BBS was found in patients with VSN, as were DAR_{AH} and DAR_{UH} (see **Figure 3**; DAR: r = –0.522, P = 0.006; DAR_{AH}: r = –0.521, P = 0.006; DAR_{UH}: r = –0.494, P = 0.01). In patients

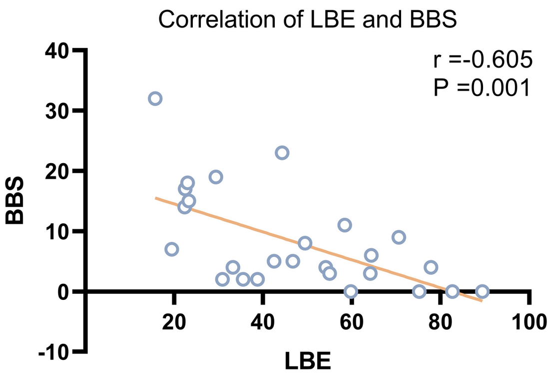


FIGURE 1
Correlation between LBE and BBS.

with VSN, no significant correlation was found between the DAR, DAR$_{\text{AH}}$, DAR$_{\text{UH}}$, BI, or FMA. In patients without VSN, no significant correlation was found between the DAR and BI, FMA, or BBS.

Pairwise-derived brain symmetry index

As shown in **Figure 4**, pdBSI values were significantly elevated in both non-VSN and VSN patients compared with healthy subjects ($F = 16.822$, $P < 0.001$; VSN vs HC: $P < 0.001$; Non-VSN vs HC: $P < 0.001$). These differences were most pronounced in the delta and theta bands. We found no significant difference in pdBSI values between patients with and without VSN. However, there were significant differences in pdBSIdelta and parieto-occipital pdBSIalpha between patients with and without VSN patients.

As shown in **Figure 5**, pdBSI showed a trend toward a negative association with FMA in stroke patients ($r = -0.508$, $P < 0.001$), mainly in the delta and theta frequency bands. No correlation was found between pdBSI and BI or BBS in both VSN and non-VSN patients.

Topographic plots

We found increased delta and theta power and decreased alpha and beta power of the EEG in VSN subjects. Regarding delta power, significant abnormalities were found in the right frontal, parietal and temporal areas in VSN patients compared to HCs, but they were located only in the right parietal areas compared to non-VSN patients. For theta power, the significantly abnormal brain regions in patients with VSN were mainly the frontal and parietal areas compared to non-VSN patients. For alpha power, significant abnormalities were found in the frontal, parietal, and occipital regions of patients with VSN compared with HCs, but they were located only in the right parietal and occipital areas compared with non-VSN patients. The topographic plots of the delta, theta, alpha, and beta frequency bands of the groups are shown in **Figure 6**.

Association between electroencephalography parameters and symptom of visuospatial neglect

The DAR showed a trend toward a positive association with LBE ($r = 0.458$, $P = 0.019$). Furthermore, the DAR$_{\text{AH}}$ showed a positive correlation with LBE ($r = 0.483$, $P = 0.012$) and ORS ($r = 0.428$, $P = 0.029$; see **Figure 7**). No correlation was found between the DAR and ORL. Parieto-occipital pdBSI$_{\text{alpha}}$ showed a trend toward a positive association with the LBE ($r = 0.442$, $P = 0.024$), but no correlation was found with the ORL or ORS (see **Figure 8**). There was no significant correlation between the pdBSI in the other frequency bands and the paper-pencil test.

Receiver operating characteristic analysis for diagnostic discrimination

The area under the ROC curve (AUC), cutoff value, sensitivity, specificity, PPV, and NPV of each EEG parameter

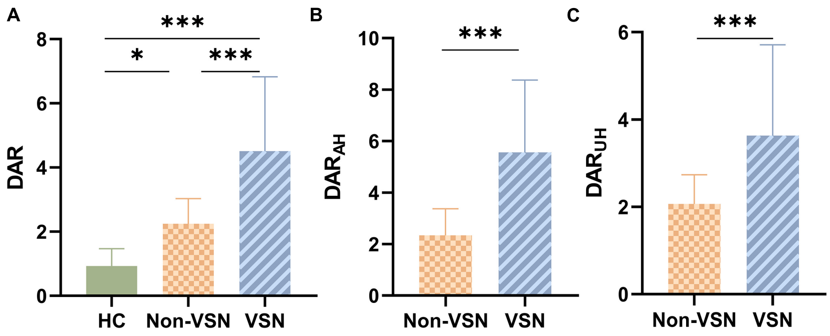


FIGURE 2
DAR values for each group. **(A)** DAR values between three groups. **(B)** DAR$_{\text{AH}}$ values between Non-VSN and VSN. **(C)** DAR$_{\text{UH}}$ values between Non-VSN and VSN.$^{*}p < 0.05$; $^{***}p < 0.001$.

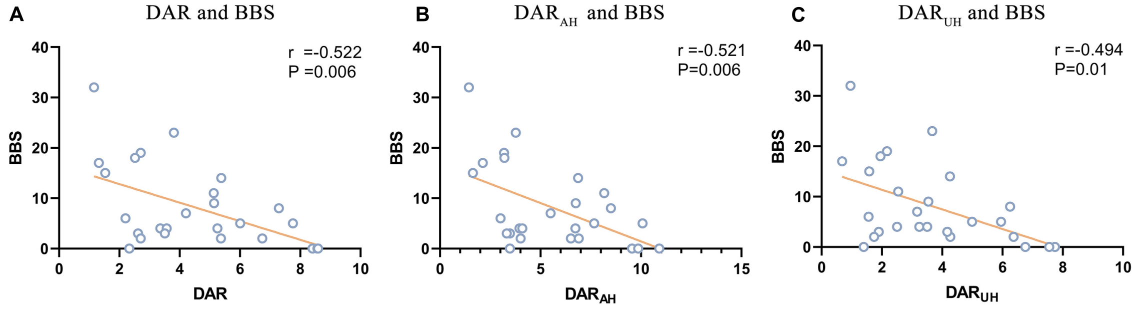


FIGURE 3
Signifcant correlations between DAR values and BBS. **(A)** Correlation between DAR and BBS. **(B)** Correlation between DAR$_{AH}$ and BBS. **(C)** Correlation between DAR$_{UH}$ and BBS.

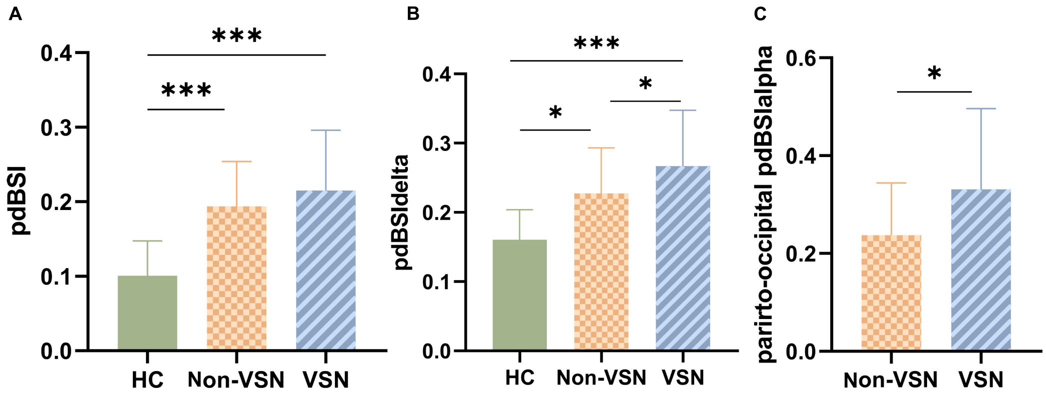


FIGURE 4
pdBSI values for each group. **(A)** pdBSI values between three groups. **(B)** pdBSIdelta values between three groups. **(C)** Parieto-occipital pdBSIalpha values between Non-VSN and VSN.**p* < 0.05; ****p* < 0.001.

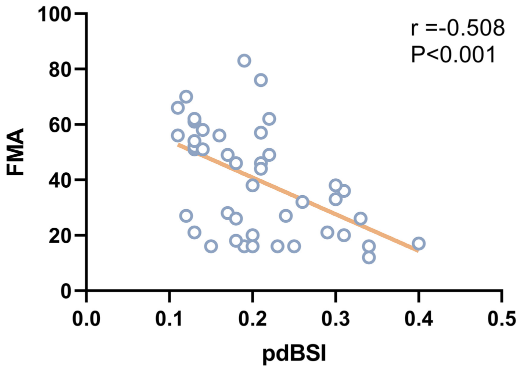


FIGURE 5
Correlation between LBE and BBS.

for differentiating between non-VSN patients and HCs were shown in Table 3. The results of ROC analyses showed that the DAR (AUC = 0.870, cut off = 1.778, $P < 0.001$), pdBSI (AUC = 0.867, cut off = 0.118, $P < 0.001$) and pdBSI$_{delta}$ (AUC = 0.728, cut off = 0.201, $P = 0.02$) could discriminate non-VSN patients and HCs. The sensitivity of the pdBSI was the highest (0.90); however, the specificity was much lower (0.63). The specificity of the DAR was the highest (0.99), while the sensitivity was 0.71.

As shown in Table 4, the ROC analysis indicated that the DAR (AUC = 0.803, cut off = 3.472, $P = 0.001$), DAR$_{AH}$ (AUC = 0.865, cut off = 3.112, $P < 0.001$), and DAR$_{UH}$ (AUC = 0.731, cut off = 3.141, $P = 0.01$) could discriminate between patients with and without VSN. The pdBSI$_{delta}$ and parieto-occipital pdBSI$_{alpha}$ had poor diagnostic ability. The sensitivity of DAR$_{AH}$ was 0.85, and the specificity was 0.78. The specificities of DAR and DAR$_{UH}$ were high (0.94 and 0.99, respectively); however, the sensitivities were generally much lower (0.62 and 0.58, respectively).

Discussion

In this study, we analyzed the resting-state EEG of three groups (patients with VSN, patients without VSN, and

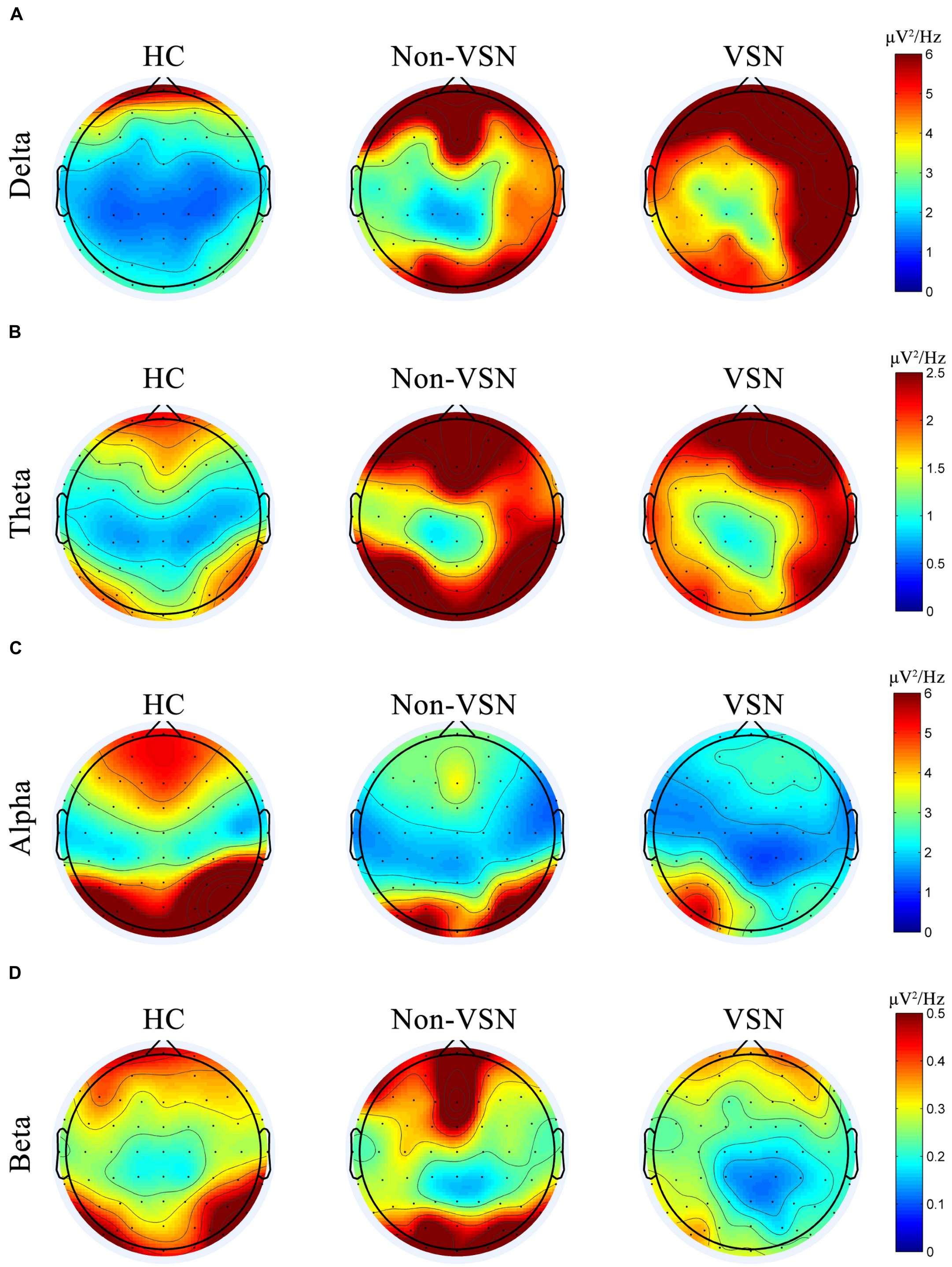


FIGURE 6
Topographic plots of the delta **(A)**, theta **(B)**, alpha **(C)**, and beta **(D)** frequency bands in three groups.

HCs) using spectral analysis. We found that patients with VSN performed poorly compared with those without VSN. Resting-state EEG has high temporal resolution, sensitivity, and specificity for evaluating the severity of VSN. Spectral analysis of the rsEEG revealed increased DAR, DAR_{AH}, DAR_{UH} and parieto-occipital $pdBSI_{alpha}$ in stroke patients with VSN, and these parameters measure the severity of VSN and are reliable markers of VSN after stroke. The DAR_{AH} was shown

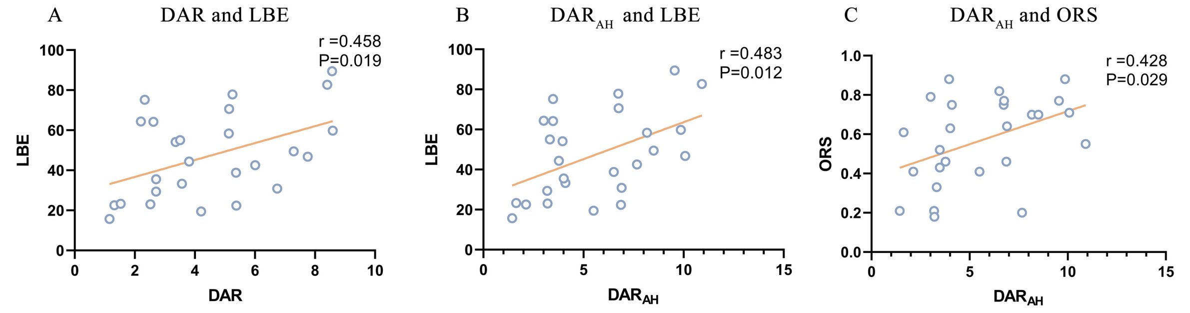


FIGURE 7
Significant correlations between DAR values and paper-pencil test. **(A)** Correlation between DAR and BBS. **(B)** Correlation between DAR$_{AH}$ and BBS. **(C)** Correlation between DAR$_{UH}$ and BBS.

to be highly sensitive to VSN and showed a good ability to discriminate poststroke patients with VSN. It has the potential to assist in the identification of VSN after stroke and is useful for clinical rehabilitation.

VSN has profound implications for quality of life after stroke; however, there is a lack of consensus regarding the screening and diagnosis of this syndrome due to the heterogeneity of its clinical manifestations. In routine stroke unit assessment, VSN was greatly underdiagnosed with a missed diagnosis rate of up to 56% (Puig-Pijoan et al., 2018). The line bisection task, line cancellation task, and star cancellation task were shown to be more sensitive tests for diagnosing VSN (Azouvi et al., 2006). However, these paper-and-pencil tests have some limitations. When ceiling and floor effects cause the scale to not truly reflect the severity of VSN, the recording of patients' electrical brain activity may be more sensitive. Because many stroke patients cannot complete task-state EEG owing to factors such as reduced cognitive level and fatigue, resting-state EEG is more suitable for VSN patients after stroke. Previous studies in healthy participants and stroke patients have shown that quantitative parameters of resting-state brain activity, such as the spectral power of different bands, are intra-individually stable in repeated measurements (Dalton et al., 2021; Duan et al., 2021). The characterization of resting-state brain activity is a reliable biomarker that may aid in clinical decision-making and treatment selection (Saes et al., 2019; Sebastian-Romagosa et al., 2020). In the current study, we explored whether spontaneous brain activity could be used as a diagnostic and assessment tool in poststroke patients with VSN.

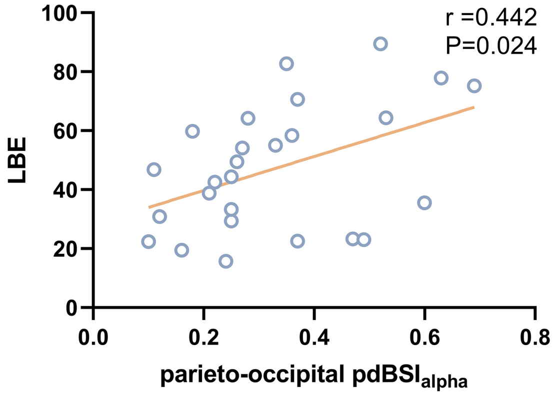


FIGURE 8
Correlation between parieto-occipital pdBSIalpha and LBE.

Higher delta and theta activity in the right fronto-parietal region and lower alpha activity in the right parieto-occipital region were found in patients with VSN. Compared with HCs, DAR values were higher in both VSN and non-VSN patients; most importantly, they were significantly higher in VSN patients than in non-VSN patients. Furthermore, the DAR and DAR$_{AH}$ were positively correlated with the paper-pencil test scale. This reveals that the DAR contains unique information regarding visuospatial neglect impairments. Excessive delta power after stroke is associated with cognitive function. Delta frequencies may reflect alertness modulation involving the corticothalamic and corticocortical neural circuits (Rabiller et al., 2015). Alpha oscillations are considered markers of vigilance, attention, cognitive processing, and cortical communication in both healthy individuals and patients (Sadaghiani and Kleinschmidt, 2016; Clayton et al., 2018). Resting EEG studies of healthy people found that alpha power in the right hemisphere was greater than that in the left hemisphere (Cicek et al., 2003). This greater right hemisphere EEG alpha activity may explain the prominent role of the right hemisphere in attention. In stroke patients with VSN, however, we found a reduction in alpha power in the right parieto-occipital region. This suggests that alpha neural oscillations may underlie the electrophysiological underpinnings of widespread attentional network connectivity in both hemispheres. VSN was initially thought to be a parietal syndrome; however, an increasing number of functional magnetic resonance imaging and EEG studies have confirmed that VSN is a disturbance in the attention network (Corbetta and Shulman, 2011; Ros et al.,

TABLE 3 Results from receiver operating characteristic (ROC) analysis to distinguish non-VSN from HC.

EEG measure	AUC [95% CI]	Cutoff value	Sensitivity	Specificity	PPV (%)	NPV (%)
DAR	0.870 [0.753, 0.987]	1.778	0.63	0.99	98.44	72.90
pdBSI	0.867 [0.751, 0.983]	0.118	0.90	0.71	75.27	87.05
$\text{pdBSI}_{\text{delta}}$	0.728 [0.562, 0.893]	0.201	0.53	0.88	81.68	65.04

AUC, area under the receiver operating curve; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value.

TABLE 4 Results from receiver operating characteristic (ROC) analysis to distinguish VSN from non-VSN.

EEG measure	AUC [95% CI]	Cutoff value	Sensitivity	Specificity	PPV (%)	NPV (%)
DAR	0.803 [0.674, 0.933]	3.472	0.62	0.94	91.65	71.03
DAR_{AH}	0.865 [0.760, 0.971]	3.112	0.85	0.78	79.21	83.48
DAR_{UH}	0.731 [0.581, 0.880]	3.141	0.58	0.99	98.29	70.06
$\text{pdBSI}_{\text{delta}}$	0.688 [0.530, 0.846]	0.252	0.62	0.78	73.48	66.90
parieto-occipital $\text{pdBSI}_{\text{alpha}}$	0.658 [0.495, 0.821]	0.212	0.81	0.50	61.77	72.25

AUC, area under the receiver operating curve; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value.

2022). The parietal lobe, particularly the posterior parietal cortex (PPC), is a critical component of the attentional network. The bilateral parietal lobes compete to mediate direct attention to the contralateral space. The posterior rsEEG alpha in healthy individuals was associated with LBT performance (Cicek et al., 2003). Task EEG studies in healthy humans have shown a relative reduction in alpha-band activity in the parieto-occipital hemisphere contralateral to the direction of spatial attention, possibly reflecting enhanced cortical excitability (Banerjee et al., 2011). In contrast, patients with VSN show pathologically enhanced alpha oscillations during both baseline fixation and cue orientation when completing a spatial orientation task (Lasaponara et al., 2019). This pathological enhancement was significantly associated with the severity of VSN and damage to white matter fiber tracts. Transcranial magnetic stimulation, inhibits cortical activity in the right PPC, disrupts attentional processes, affects visuospatial attention, and induces transient spatial neglect-like symptoms in healthy adults (Fierro et al., 2000; Mariner et al., 2021). Moreover, multiple studies have demonstrated that transcranial magnetic stimulation of the PPC can effectively improve the symptoms of patients with VSN after stroke (Salazar et al., 2018; Ye et al., 2021). Therefore, alpha oscillations in the parietal cortex are a reliable biomarker of visuospatial neglect after stroke and may be useful for rehabilitation interventions involving non-invasive brain stimulation and EEG-based neurofeedback.

Not only did DAR values increase in the affected hemisphere after stroke compared to healthy individuals, but DAR values also increased in the unaffected hemisphere. Although structural damage from stroke is focal, remote dysfunction may occur in areas of the brain that are distant from the damaged area (Siegel et al., 2016). This view of distributed brain network connectivity disturbances provides new insights into the recovery from post-stroke dysfunction. The bimodal balance recovery theory suggests that the recovery of dysfunction in some stroke patients may be related to the contralateral hemisphere (Di Pino et al., 2014), and this has been confirmed in patients with visuospatial neglect (Cao et al., 2016). In fact, the DAR_{AH} was higher than the DAR_{UH} in some patients with VSN, and the opposite was observed in other patients in our study. Whether this inconsistency is related to the patient's recovery pattern is unclear and further research is warranted.

Analyses of the pdBSI index indicated no significant hemispheric asymmetry in healthy participants. In contrast, patients with damage to the right hemisphere showed significantly increased low-frequency oscillations in the lesional hemisphere. Our results showed that patients with subacute right hemisphere stroke had significantly higher pdBSI values than healthy subjects. Higher BSI values reflect a greater power asymmetry in the hemispheres. This finding is consistent with those of other studies (Sheorajpanday et al., 2011). However, we found no significant difference in pdBSI values between patients with and without VSN. Indeed, previous studies have found that pdBSI is significantly associated with infarct volume after controlling for various confounding factors (Sheorajpanday et al., 2011). The interhemispheric asymmetry represented by pdBSI persisted during the chronic phase of stroke (Saes et al., 2019). In healthy individuals, the bilateral hemispheres competitively inhibit each other to achieve balance, but this balance is broken after a stroke. The inhibitory effect of the contralateral hemisphere on the affected hemisphere was higher, which results in interhemispheric asymmetry. Interhemispheric asymmetry can affect recovery from dysfunction. Our study found that higher pdBSI values, mainly $\text{pdBSI}_{\text{delta}}$ and $\text{pdBSI}_{\text{theta}}$, but not DAR, were associated with lower FMA in patients with stroke with or without VSN, which is consistent with the results of some studies (Saes et al., 2019). However, other studies have come to the opposite conclusion that compared with differences between cerebral hemispheres, DAR values can be more sensitive in assessing the severity of dyskinesia in stroke patients (Brito et al., 2021). The low number of EEG channels (nine scalp electrodes) used in the study by Brito et al. may have contributed to these inconsistent results. The slower delta and theta frequencies are thought to be generated by cortical layers II-VI. Low-frequency cortical activity may reflect the integrity of the cortical-cortical network connectivity. Previous studies also found that higher $\mathrm{BSI}_{\mathrm{theta}}$ values were significantly negatively correlated with upper extremity motor function 6 months after stroke (Saes et al., 2020, 2021). Therefore, low-frequency oscillations may reflect both injury and recovery after stroke and may be a reliable biomarker for stroke rehabilitation (Cassidy et al., 2020). Our study found that visuospatial neglect following stroke affects the EEG alpha rhythm, mainly in the right parietal and occipital areas. This resulted in marked asymmetry of the alpha band in the parietal and occipital regions, which was significantly associated with neglect severity. These results showed that interhemispheric asymmetry in the alpha band of the parieto-occipital region can provide a measure of the severity of neglect.

Thus, rsEEG may be a useful tool for identifying patients with VSN after stroke. In the present study, we found that the use of the $\mathrm{DAR}_{\mathrm{AH}}$ to distinguish between patients with VSN and non-VSN patients was more than 80% sensitive and 70% specific. Therefore, $\mathrm{DAR}_{\mathrm{AH}}$ is a promising marker for the diagnosis of VSN after stroke. This finding is important for the early diagnosis of VSN after stroke. In distinguishing non-VSN patients from healthy individuals, DAR is an indicator with high specificity and low sensitivity, while pdBSI is an indicator with high sensitivity and low specificity. MRI is contraindicated in some patients and expensive. In contrast, EEG may allow an inexpensive, reliable bedside evaluation with practically no contraindications. Due to the advantages of EEG, there have been many studies on EEG-assisted diagnosis of acute stroke (Erani et al., 2020; van Meenen et al., 2021). The low sensitivity of the DAR in this study was inconsistent with that of previous studies. Studies have found that a DAR of 3.7 has 100% sensitivity and 100% specificity in distinguishing acute ischemic stroke from healthy individuals (Finnigan et al., 2016). This study collected patients in the acute phase, while our study collected patients in the subacute phase. The spectral signature of rsEEG changes over time, especially in the delta frequency band. Differences in EEG acquisition time and processing methods may have contributed to the differences in the studies. Therefore, to facilitate the use of rsEEG for stroke diagnosis, standardized EEG acquisition and processing procedures are required. In addition to this, larger sample size and other methods, such as coherence, should also be used to assess the accuracy of rsEEG aids in diagnosis.

Some studies found that VSN patients were more dependent on ADL than non-VSN patients (Bosma et al., 2020), a finding that is consistent with our results. VSN can disrupt a patient's balance and affect motor function recovery. Compared with the line bisection and star cancellation tasks, no correlation was found between the line cancellation task and the resting-state EEG parameters. This may be because the line cancellation task contained only 30 line segments, and the calculated omission rate was not as sensitive as that of the line bisection task or star cancellation task.

The current study had several limitations. First, the sample size of the current study is insufficient, which weakens the influence of the article; however, we will conduct further studies with a larger sample size to confirm the feasibility of the research conclusions. Second, we only assessed patients at admission and did not follow up with these patients. Although our study demonstrated that resting-state EEG may be a useful tool for identifying potential VSN after stroke, EEG parameters can also reflect the severity of VSN after stroke. However, over time, the spectral features show a gradual normalization. Studies have found that the DAR value of patients with chronic stroke is not different from that of healthy people. We do not know whether rsEEG remains a useful assessment tool for VSN after spontaneous recovery in patients with VSN. A longitudinal assessment of patients with stroke was not performed, and the relationship between EEG parameters and VSN could not be clearly defined. Future studies should expand the single time point to multiple time points to verify our conclusions further. Third, stroke was divided into acute, subacute, and chronic phases, but this study only included patients in the subacute phase; therefore, further research is needed to perform subgroup analysis on acute and chronic phase patients. Fourth, due to lack of MRI data, this study only corrected for lesion location without calculating lesion volume. However, it did not calculate the lesion area in stroke patients.

Conclusion

To the best of our knowledge, this is the first study on VSN resting-state EEG spectral analysis after stroke. Overall, our EEG results are consistent with those of previous EEG studies and provide new evidence for rsEEG features of visuospatial neglect after stroke. The rsEEG cortical asymmetry and the DAR were increased in patients with visuospatial neglect after stroke. Higher asymmetry in the parieto-occipital region of the alpha band and higher DAR values are associated with more severe visuospatial neglect. Furthermore, our study showed that resting-state DARAH can accurately differentiate between patients with and without VSN after stroke. This implies that rsEEG could be used for the auxiliary diagnosis of VSN after stroke, and the DAR and pdBSI alpha in resting EEG may be useful biomarkers of visuospatial neglect after stroke.

Data availability statement

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Ethics statement

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

Author contributions

YZ: drafting and revising the manuscript, study concept and design, analysis and interpretation of data. LY: study concept and design, acquisition of data, revising the manuscript, and obtaining funds. LC: study concept and design, and revising the manuscript. WS: study concept and design, analysis and interpretation of data, and study supervision. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (Grant Number 82002386) and the Beijing Municipal Administration of Hospitals Incubating Program (code: PX2021037). The funding sources had no involvement in the study design, in the collection, analysis and interpretation of data, in the writing of the report, or in the decision to submit this article for publication.

Acknowledgments

We would like to thank all our participants for their interest and time investment.

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

References

Aravind, G., and Lamontagne, A. (2018). Effect of visuospatial neglect on spatial navigation and heading after stroke. *Ann. Phys. Rehabil. Med.* 61, 197–206. doi: 10.1016/j.rehab.2017.05.002

Azouvi, P. (2017). The ecological assessment of unilateral neglect. *Ann. Phys. Rehabil. Med.* 60, 186–190. doi: 10.1016/j.rehab.2015.12.005

Azouvi, P., Bartolomeo, P., Beis, J. M., Perennou, D., Pradat-Diehl, P., and Rousseaux, M. (2006). A battery of tests for the quantitative assessment of unilateral neglect. *Restor. Neurol. Neurosci.* 24, 273–285.

Banerjee, S., Snyder, A. C., Molholm, S., and Foxe, J. J. (2011). Oscillatory alpha-band mechanisms and the deployment of spatial attention to anticipated auditory and visual target locations: Supramodal or sensory-specific control mechanisms? *J. Neurosci.* 31, 9923–9932. doi: 10.1523/JNEUROSCI.4660-10.2011

Bosma, M. S., Nijboer, T. C. W., Caljouw, M. A. A., and Achterberg, W. P. (2020). Impact of visuospatial neglect post-stroke on daily activities, participation and informal caregiver burden: A systematic review. *Ann. Phys. Rehabil. Med.* 63, 344–358. doi: 10.1016/j.rehab.2019.05.006

Brito, R., Baltar, A., Berenguer-Rocha, M., Shirahige, L., Rocha, S., Fonseca, A., et al. (2021). Intrahemispheric EEG: A New Perspective for Quantitative EEG Assessment in Poststroke Individuals. *Neural Plast.* 2021:5664647. doi: 10.1155/2021/5664647

Cao, L., Fu, W., Zhang, Y., Huo, S., Du, J., Zhu, L., et al. (2016). Intermittent theta burst stimulation modulates resting-state functional connectivity in the attention network and promotes behavioral recovery in patients with visual spatial neglect. *Neuroreport* 27, 1261–1265. doi: 10.1097/WNR.0000000000000689

Cassidy, J. M., Wodeyar, A., Wu, J., Kaur, K., Masuda, A. K., Srinivasan, R., et al. (2020). Low-Frequency Oscillations Are a Biomarker of Injury and Recovery After Stroke. *Stroke* 51, 1442–1450. doi: 10.1161/STROKEAHA.120.028932

Chen, P., Hreha, K., Kong, Y., and Barrett, A. M. (2015). Impact of spatial neglect on stroke rehabilitation: Evidence from the setting of an inpatient rehabilitation facility. *Arch. Phys. Med. Rehabil.* 96, 1458–1466. doi: 10.1016/j.apmr.2015.03.019

Cicek, M., Nalcaci, E., and Kalaycioglu, C. (2003). Line bisection task performance and resting EEG alpha power. *Int. J. Neurosci.* 113, 849–866. doi: 10.1080/00207450390200981

Clayton, M. S., Yeung, N., and Cohen Kadosh, R. (2018). The many characters of visual alpha oscillations. *Eur. J. Neurosci.* 48, 2498–2508. doi: 10.1111/ejn.13747

Corbetta, M., and Shulman, G. L. (2011). Spatial neglect and attention networks. *Annu. Rev. Neurosci.* 34, 569–599. doi: 10.1146/annurev-neuro-061010-113731

Dalton, S. G. H., Cavanagh, J. F., and Richardson, J. D. (2021). Spectral Resting-State EEG (rsEEG) in Chronic Aphasia Is Reliable, Sensitive, and Correlates With Functional Behavior. *Front. Hum. Neurosci.* 15:624660. doi: 10.3389/fnhum.2021.624660

Di Pino, G., Pellegrino, G., Assenza, G., Capone, F., Ferreri, F., Formica, D., et al. (2014). Modulation of brain plasticity in stroke: A novel model for neurorehabilitation. *Nat. Rev. Neurol.* 10, 597–608. doi: 10.1038/nrneurol.2014.162

Duan, W., Chen, X., Wang, Y. J., Zhao, W., Yuan, H., and Lei, X. (2021). Reproducibility of power spectrum, functional connectivity and network construction in resting-state EEG. *J. Neurosci. Methods* 348:108985. doi: 10.1016/j.jneumeth.2020.108985

Erani, F., Zolotova, N., Vanderschelden, B., Khoshab, N., Sarian, H., Nazarzai, L., et al. (2020). Electroencephalography Might Improve Diagnosis of Acute Stroke and Large Vessel Occlusion. *Stroke* 51, 3361–3365. doi: 10.1161/STROKEAHA.120.030150

Esposito, E., Shekhtman, G., and Chen, P. (2020). Prevalence of spatial neglect post-stroke: A systematic review. *Ann. Phys. Rehabil. Med.* 64:101459. doi: 10.1016/j.rehab.2020.10.010

Fierro, B., Brighina, F., Oliveri, M., Piazza, A., La Bua, V., Buffa, D., et al. (2000). Contralateral neglect induced by right posterior parietal rTMS in healthy subjects. *Neuroreport* 11, 1519–1521.

Fingelkurts, A. A., and Fingelkurts, A. A. (2015). Altered structure of dynamic electroencephalogram oscillatory pattern in major depression. *Biol. Psychiatry* 77, 1050–1060. doi: 10.1016/j.biopsych.2014.12.011

Finnigan, S., Wong, A., and Read, S. (2016). Defining abnormal slow EEG activity in acute ischaemic stroke: Delta/alpha ratio as an optimal QEEG index. *Clin. Neurophysiol.* 127, 1452–1459. doi: 10.1016/j.clinph.2015.07.014

Hammerbeck, U., Gittins, M., Vail, A., Paley, L., Tyson, S. F., and Bowen, A. (2019). Spatial Neglect in Stroke: Identification, Disease Process and Association with Outcome During Inpatient Rehabilitation. *Brain Sci.* 9:374. doi: 10.3390/brainsci9120374

Heilman, K. M., and Valenstein, E. (1979). Mechanisms underlying hemispatial neglect. *Ann. Neurol.* 5, 166–170. doi: 10.1002/ana.410050210

Herrmann, C. S., Struber, D., Helfrich, R. F., and Engel, A. K. (2016). EEG oscillations: From correlation to causality. *Int. J. Psychophysiol.* 103, 12–21. doi: 10.1016/j.ijpsycho.2015.02.003

Jehkonen, M., Ahonen, J. P., Dastidar, P., Koivisto, A. M., Laippala, P., Vilkki, J., et al. (2000). Visual neglect as a predictor of functional outcome one year after stroke. *Acta Neurol. Scand.* 101, 195–201. doi: 10.1034/j.1600-0404.2000.101003195.x

Lasaponara, S., Pinto, M., Aiello, M., Tomaiuolo, F., and Doricchi, F. (2019). The Hemispheric Distribution of alpha-Band EEG Activity During Orienting of Attention in Patients with Reduced Awareness of the Left Side of Space (Spatial Neglect). *J. Neurosci.* 39, 4332–4343. doi: 10.1523/JNEUROSCI.2206-18.2019

Mariner, J., Loetscher, T., and Hordacre, B. (2021). Parietal Cortex Connectivity as a Marker of Shift in Spatial Attention Following Continuous Theta Burst Stimulation. *Front. Hum. Neurosci.* 15:718662. doi: 10.3389/fnhum.2021.718662

Nijboer, T., van de Port, I., Schepers, V., Post, M., and Visser-Meily, A. (2013). Predicting functional outcome after stroke: The influence of neglect on basic activities in daily living. *Front. Hum. Neurosci.* 7:182. doi: 10.3389/fnhum.2013.00182

Parton, A., Malhotra, P., and Husain, M. (2004). Hemispatial neglect. *J. Neurol. Neurosurg. Psychiatry* 75, 13–21.

Pirondini, E., Goldshuv-Ezra, N., Zinger, N., Britz, J., Soroker, N., Deouell, L. Y., et al. (2020). Resting-state EEG topographies: Reliable and sensitive signatures of unilateral spatial neglect. *Neuroimage Clin.* 26:102237. doi: 10.1016/j.nicl.2020.102237

Puig-Pijoan, A., Giralt-Steinhauer, E., Zabalza, de Torres, A., Manero Borras, R. M., Sanchez-Benavides, G., et al. (2018). Underdiagnosis of Unilateral Spatial Neglect in stroke unit. *Acta Neurol. Scand.* 138, 441–446. doi: 10.1111/ane.12998

Rabiller, G., He, J. W., Nishijima, Y., Wong, A., and Liu, J. (2015). Perturbation of Brain Oscillations after Ischemic Stroke: A Potential Biomarker for Post-Stroke Function and Therapy. *Int. J. Mol. Sci.* 16, 25605–25640. doi: 10.3390/ijms161025605

Ros, T., Michela, A., Mayer, A., Bellmann, A., Vuadens, P., Zermatten, V., et al. (2022). Disruption of large-scale electrophysiological networks in stroke patients with visuospatial neglect. *Netw. Neurosci.* 6, 69–89. doi: 10.1162/netn_a_00210

Sadaghiani, S., and Kleinschmidt, A. (2016). Brain Networks and alpha-Oscillations: Structural and Functional Foundations of Cognitive Control. *Trends Cogn. Sci.* 20, 805–817. doi: 10.1016/j.tics.2016.09.004

Saes, M., Meskers, C. G. M., Daffertshofer, A., de Munck, J. C., Kwakkel, G., van Wegen, E. E. H., et al. (2019). How does upper extremity Fugl-Meyer motor score relate to resting-state EEG in chronic stroke? A power spectral density analysis. *Clin. Neurophysiol.* 130, 856–862. doi: 10.1016/j.clinph.2019.01.007

Saes, M., Meskers, C. G. M., Daffertshofer, A., van Wegen, E. E. H., Kwakkel, G., and consortium, D. E. (2021). Are early measured resting-state EEG parameters predictive for upper limb motor impairment six months poststroke? *Clin. Neurophysiol.* 132, 56–62. doi: 10.1016/j.clinph.2020.09.031

Saes, M., Zandvliet, S. B., Andringa, A. S., Daffertshofer, A., Twisk, J. W. R., Meskers, C. G. M., et al. (2020). Is Resting-State EEG Longitudinally Associated With Recovery of Clinical Neurological Impairments Early Poststroke? A Prospective Cohort Study. *Neurorehabil. Neural Repair.* 34, 389–402. doi: 10.1177/1545968320905797

Salazar, A. P. S., Vaz, P. G., Marchese, R. R., Stein, C., Pinto, C., and Pagnussat, A. S. (2018). Noninvasive Brain Stimulation Improves Hemispatial Neglect After Stroke: A Systematic Review and Meta-Analysis. *Arch. Phys. Med. Rehabil.* 99, 355–366.e351. doi: 10.1016/j.apmr.2017.07.009

Sebastian-Romagosa, M., Udina, E., Ortner, R., Dinares-Ferran, J., Cho, W., Murovec, N., et al. (2020). EEG Biomarkers Related With the Functional State of Stroke Patients. *Front. Neurosci.* 14:582. doi: 10.3389/fnins.2020.00582

Sheorajpanday, R. V., Nagels, G., Weeren, A. J., van Putten, M. J., and De Deyn, P. P. (2009). Reproducibility and clinical relevance of quantitative EEG parameters in cerebral ischemia: A basic approach. *Clin. Neurophysiol.* 120, 845–855. doi: 10.1016/j.clinph.2009.02.171

Sheorajpanday, R. V., Nagels, G., Weeren, A. J., van Putten, M. J., and De Deyn, P. P. (2011). Quantitative EEG in ischemic stroke: Correlation with functional status after 6 months. *Clin. Neurophysiol.* 122, 874–883. doi: 10.1016/j.clinph.2010.07.028

Siegel, J. S., Ramsey, L. E., Snyder, A. Z., Metcalf, N. V., Chacko, R. V., Weinberger, K., et al. (2016). Disruptions of network connectivity predict impairment in multiple behavioral domains after stroke. *Proc. Natl. Acad. Sci. U.S.A.* 113:E4367–E4376. doi: 10.1073/pnas.1521083113

Ten Brink, A. F., Verwer, J. H., Biesbroek, J. M., Visser-Meily, J. M. A., and Nijboer, T. C. W. (2017). Differences between left- and right-sided neglect revisited: A large cohort study across multiple domains. *J. Clin. Exp. Neuropsychol.* 39, 707–723. doi: 10.1080/13803395.2016.1262333

van Meenen, L. C. C., van Stigt, M. N., Siegers, A., Smeekes, M. D., van Grondelle, J. A. F., Geuzebroek, G., et al. (2021). Detection of Large Vessel Occlusion Stroke in the Prehospital Setting: Electroencephalography as a Potential Triage Instrument. *Stroke* 52:e347–e355. doi: 10.1161/STROKEAHA.120.033053

Ye, L. L., Xie, H. X., Cao, L., and Song, W. Q. (2021). Therapeutic Effects of Transcranial Magnetic Stimulation on Visuospatial Neglect Revealed With Event-Related Potentials. *Front. Neurol.* 12:799058. doi: 10.3389/fneur.2021.799058

Zappasodi, F., Tecchio, F., Marzetti, L., Pizzella, V., Di Lazzaro, V., and Assenza, G. (2019). Longitudinal quantitative electroencephalographic study in mono-hemispheric stroke patients. *Neural Regen. Res.* 14, 1237–1246. doi: 10.4103/1673-5374.251331

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 16 August 2022
DOI 10.3389/fnins.2022.975131

Unsupervised layer-wise feature extraction algorithm for surface electromyography based on information theory

Mingqiang Li[1]*†, Ziwen Liu[2]†, Siqi Tang[1], Jianjun Ge[1] and Feng Zhang[1]

[1]Information Science Academy, China Electronics Technology Group Corporation, Beijing, China, [2]School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing, China

EDITED BY
Xiaomao Fan,
South China Normal University, China

REVIEWED BY
Tiexiang Wen,
Shenzhen Technology University,
China
Yun Li,
Sun Yat-sen University, China
Xianhui Chen,
New York University, United States

*CORRESPONDENCE
Mingqiang Li
limingqiang14@mails.ucas.ac.cn

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

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

RECEIVED 21 June 2022
ACCEPTED 25 July 2022
PUBLISHED 16 August 2022

CITATION
Li M, Liu Z, Tang S, Ge J and Zhang F (2022) Unsupervised layer-wise feature extraction algorithm for surface electromyography based on information theory.
Front. Neurosci. 16:975131.
doi: 10.3389/fnins.2022.975131

Feature extraction is a key task in the processing of surface electromyography (SEMG) signals. Currently, most of the approaches tend to extract features with deep learning methods, and show great performance. And with the development of deep learning, in which supervised learning is limited by the excessive expense incurred due to the reliance on labels. Therefore, unsupervised methods are gaining more and more attention. In this study, to better understand the different attribute information in the signal data, we propose an information-based method to learn disentangled feature representation of SEMG signals in an unsupervised manner, named Layer-wise Feature Extraction Algorithm (LFEA). Furthermore, due to the difference in the level of attribute abstraction, we specifically designed the layer-wise network structure. In TC score and MIG metric, our method shows the best performance in disentanglement, which is 6.2 lower and 0.11 higher than the second place, respectively. And LFEA also get at least 5.8% accuracy lead than other models in classifying motions. All experiments demonstrate the effectiveness of LEFA.

KEYWORDS

information theory, feature extraction, unsupervised learning, information bottleneck, disentangled representation, surface electromyography

Introduction

Feature engineering is an important component of pattern recognition and signal processing. Learning good representations from observed data can help reveal the underlying structures. In recent decades, feature extraction methods (He et al., 2016; Howard et al., 2017; Hassani and Khasahmadi, 2020; Zbontar et al., 2021) have drawn considerable attention. Due to the high cost of obtaining labels, supervised learning methods suffer from data volume limitations. Unsupervised learning methods therefore becomes critical for feature extraction. Most of these are based on probabilistic models, such as maximum likelihood estimation (Myung, 2003), maximum *a posteriori* probability estimation (Richard and Lippmann, 1991), and mutual information (MI) (Thomas and Joy, 2006). Methods such as principal component analysis (PCA) (Abdi and Williams, 2010), linear discriminant analysis (Izenman, 2013), isometric feature mapping (Tenenbaum et al., 2000), and Laplacian eigenmaps (Belkin and Niyogi, 2003) are widely used owing to their good performance, high efficiency, flexibility, and simplicity. Other algorithms are based on reconstruction errors or generative criteria, such as autoencoders (Bengio et al., 2013) and generative adversarial networks (GANs) (Goodfellow et al., 2014). Occasionally, the reconstruction error criterion also has a probabilistic interpretation.

In recent years, deep learning has become a dominant method of representation learning, particularly in the supervised case. A neural network simulates the mechanism of hierarchical information processing in the brain and is optimized using the back propagation (BP) algorithm (LeCun et al., 1988). Because several feature engineering tasks are unsupervised, that is, no label information is available in the real situation and collecting considerable labeled data is expensive, methods to discover the feature representation in an unsupervised case have been significantly developed in recent years. MI maximization (Bell and Sejnowski, 1995) and minimization criteria (Matsuda and Yamaguchi, 2003) are powerful tools for capturing salient features of data and disentangling these features. In particular, variational autoencoder (VAE) (Kingma and Welling, 2013) based models and GAN have exhibited effective applications in disentangled representations. There are two benefits of learning disentangled representations. First, models with disentangled representations are more explainable (Bengio et al., 2013; Liu et al., 2021). Second, disentangled representations make it easier and more efficient to manipulate training-data synthesis. However, the backpropagation algorithm still requires a high amount of computation and data.

To extract features information in SEMG signal data, we propose a Layer-wise Feature Extraction Algorithm (LFEA) based on information theory in the unsupervised case, which includes a hierarchical structure to capture disentangled features. In each layer, we split the feature into two independent blocks, and ensure the information separation between the blocks *via* information constraint, which we called Information Separation Module (ISM). Moreover, to ensure the expressiveness of the representation without losing crucial information, we propose the Information Representation Module (IRM) to enable the learned representation to reconstruct the original signal data. Meanwhile, redundant information would affect the quality of the representation and thus degrade the effectiveness of downstream tasks, for which Information Compression Module (ICM) is proposed to reduce the redundant and noisy information. In terms of the optimization algorithm, our back-propagation process is only performed in a single layer and not back propagated throughout the network, which can greatly reduce the amount of computation while having no effect on the effectiveness of our method. Regarding the experiments, we have made improvement and strengths in terms of motion classification and representation disentanglement over the traditional methods of surface electromyography (SEMG). Especially, on NinaPro database 2 (DB2) dataset, our approach gets a significant 4% improvement in the motion classification, and better model stability.

This manuscript is organized as follows. In Section 2, we introduce the related work. The proposed method LFEA is described in Section 3. We present the numerical results in Section 4. Section 5 gives the conclusion of this manuscript.

Related work

Disentangled representation

The disentanglement problem has played a significant role, particularly because of its better interpretability and controllability. The VAE variants construct representations in which each dimension is independent and corresponds to a dedicated attribute. β-VAE (Higgins et al., 2016) adds a hyperparameter to control the trade-off between compression and expression. An analysis of β-VAE by Burgess et al. (2018) is provided, and the capacity term is proposed to obtain a better balance of the reconstruction error. Penalizing the total correlation term to reinforce the independence among representation dimensions was proposed in Factor VAE (Kim and Mnih, 2018) and β-TCVAE (Chen et al., 2018). FHVAE (Hsu et al., 2017) and DSVAE (Yingzhen and Mandt, 2018) constructed a new model architecture and factorized the latent variables into static and dynamic parts. Cheng et al. (2020b) described a GAN model using MI. Similar to our study, Gonzalez-Garcia et al. (2018) proposed a model to disentangle the attributes of paired data into shared and exclusive representations.

Information theory

Shannon's MI theory (Shannon, 2001) is a powerful tool for characterizing good representation. However, one major problem encountered in the practical application of information theory is computational difficulties in high-dimensional spaces. Numerous feasible computation methods have been proposed, such as Monte Carlo sampling, population coding, and the mutual information neural estimator (Belghazi et al., 2018). In addition, the information bottleneck (IB) principle

(Tishby et al., 2000; Tishby and Zaslavsky, 2015; Shwartz-Ziv and Tishby, 2017; Jeon et al., 2021) learns an informative latent representation of target attributes. A variational model to make IB computation easier was introduced in variational IB (Alemi et al., 2016). A stair disentanglement net was proposed to capture attributes in respective aligned hidden spaces and extend the IB principle to learn a compact representation.

Surface electromyography signal feature extraction

With the development of SEMG signal acquisition technology, the analysis and identification of SEMG signals has also drawn the attention of researchers.

As machine learning has demonstrated excellent feature extraction capabilities in areas such as images and speech, it can also be a good solution for recognizing SEMG signals. The basic motivation was to construct and simulate neural networks for human brain analysis and learning. Deep neural networks can extract the features of SEMG signals while effectively avoiding the absence of valid information in the signal and improving the accuracy of recognition. Xing et al. (2018) used a parallel architecture model with five convolutional neural networks to extract and classify SEMG signals. Atzori et al. (2016) used a convolutional network to classify an average of 50 hand movements from 67 intact subjects and 11 transradial amputees, achieving a better recognition accuracy than traditional machine learning methods. Zhai et al. (2017) proposed a self-calibrating classifier. This can automatically calibrate the original classifier. The calibrated classifier also obtains a higher accuracy than the uncalibrated classifier. In addition, He et al. (2018) incorporated a long short-term memory network (Hochreiter and Schmidhuber, 1997) into a multilayer perceptron and achieved better classification of SEMG signals in the NinaPro DB1 dataset.

As stated, deep learning methods can help overcome the limitations of traditional methods and lead to better performance of SEMG. Furthermore, deep-learning methods can provide an extensive choice of models to satisfy different conditional requirements.

Method

Preliminary

Information theory is commonly used to describe stochastic systems. Among the dependency measurements, mutual information (MI) was used to measure the correlation between random variables or factors. Given two random variables X and Z, the MI is defined as follows:

$$I\left(X; Z\right) = E_{p(x,z)}\left[\log \frac{p(x, z)}{p(x)p(z)}\right] \tag{1}$$

Regarding the data processing flow as a Markov chain $X \rightarrow Z \rightarrow Y$, the information bottleneck (IB) principle desires that the useful information in the input X can pass through the 'bottleneck' while the noise and irrelevant information are filtered out. The IB principle is expressed as follow:

$$min\ \mathrm{R}_{IB} = I\left(X; Z\right) - \beta I(Z; Y) \tag{2}$$

where, β is the tradeoff parameter between the complexity of the representation and the amount of relevant essential information.

Framework

The diagram of our proposed Layer-wise Feature Extraction Algorithm (LFEA) is illustrated in **Figure 1**. Our algorithm aims to learn a representation that satisfies three main properties: "Compression," "Expression" and "Disentanglement." To this end, three key information process modules are introduced, including the information compression module (ICM), information expression module (IEM), and information separation module (ISM) in each layer.

In the ICM, input s^{i-1} of layer i is compressed into h^i ($s^0 = X$). In the IEM, z^i as part of h^i is constrained to represent the original input X. In the ISM section, s^i and z^i are irrelevant. The parameters of the ICM and IEM in layer i are denoted as ϕ^i and θ^i. The data information flow can be expressed as follows:

$$h^i \sim q_{\phi^i}\left(h^i|s^{i-1}\right), \tag{3}$$

$$h^i = \left(z^i, s^i\right), \tag{4}$$

$$\tilde{X} \sim p_{\theta^i}\left(\tilde{X}|z^i\right), \tag{5}$$

where, $s^0 = X$, and q_{ϕ^i} and p_{θ^i} are the condition distributions with ϕ^i and θ^i for h^i and $\tilde{X}$. In following sections, we describe these three modules in detail.

Information compression module

According to (3), h^i is the hidden representation of s^{i-1}. To ensure information 'compression,' the optimal representation of s^{i-1} should forget redundant information altogether, that is, h^i represents s^{i-1} with the lowest bits. Formally, the objective in the i-th layer to be minimized is as follows:

$$minL_{ICM} \triangleq I_{\phi^i}\left(S^{i-1}; H^i\right) \tag{6}$$

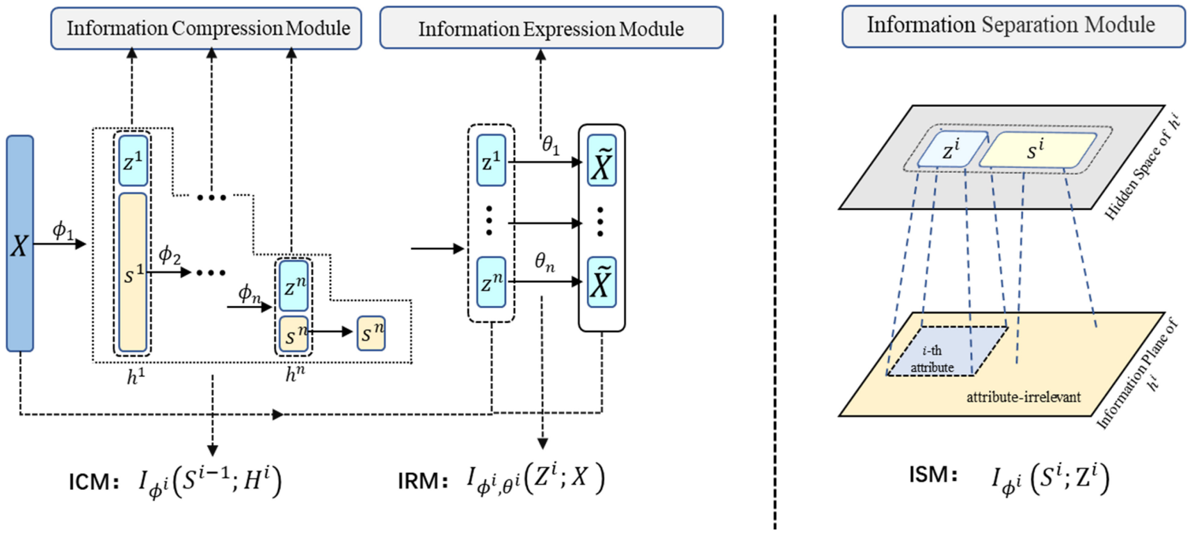


FIGURE 1
The diagram of Layer-wise Feature Extraction Algorithm (LFEA). LEFA contains three core modules: Information Compression Module (ICM), Information Expression Module (IEM) and Information Separation Module (ISM), to ensure compression, expression and disentanglement of representation, respectively.

Due to intractability of mutual information, optimizing L_{ICM} with gradient methods directly is not feasible. We therefore derived the upper bound of L_{ICM} with the variational inference method and get decomposition as follows:

$$I_{\phi^i}\left(S^{i-1};H^i\right)=E_{q_{\phi^i}\left(s^{i-1},h^i\right)}\left[\log\frac{q_{\phi^i}\left(h^i|s^{i-1}\right)p\left(h\right)}{q_{\phi^i}\left(h^i\right)p\left(h\right)}\right]$$

$$=L^{upper}_{ICM}-D_{KL}\left(q_{\phi^i}\left(h^i\right)||p\left(h\right)\right), \qquad (7)$$

where, $p\left(h\right)$ is the prior, and L^{upper}_{ICM} is the upper bound of L_{ICM} defined as follows:

$$L^{upper}_{ICM}=E_{q_{\phi^i}\left(s^{i-1}\right)}\left[D_{KL}\left(q_{\phi^i}\left(h^i|s^{i-1}\right)||p\left(h\right)\right)\right],$$

$$D_{KL}\left(P,Q\right)=\mathrm{E_P}\left[\log\frac{\mathrm{p}}{\mathrm{q}}\right]. \qquad (8)$$

Information expression module

With the ICM guaranteeing the information compression, LFEA also need to ensure the expressiveness of the representation to the data. We therefore propose the information expression module (IEM). To ensure sufficient information to reconstruct the original data X, we maximize the MI between and Z^i in i-th layer, that is,

$$\mathbf{max}\mathbf{L}_{IEM}\triangleq\mathbf{I}_{\phi^i,\theta^i}\left(\mathbf{z^i};X\right) \qquad (9)$$

For L_{IEM}, we can obtain a lower bound using the variational approximation method as follows:

$$L_{IEM}\geq L^{lower}_{IEM}-D_{KL}\left(p\left(x\right)||p_{\theta^i}\left(x\right)\right), \qquad (10)$$

where, $p_{\theta^i}\left(x\right)$

$$L^{lower}_{IEM}=E_{p(x)}\left[E_{q_{\phi^i}\left(z^i|x\right)}\log p_{\theta^i}\left(x|z^i\right)\right] \qquad (11)$$

can be viewed as the reconstruction loss.

Information separation module

To achieve disentanglement of representations (Independent of each block $z^1,z^2,\ldots,z^n$ in Z), we further introduce the information separation module (ISM) in each layer. In i-th layer, the principle of ISM is to ensure that there is no intersection information between z^i and s^i, that is,

$$\mathbf{max}\mathbf{L}_{ISM}\triangleq\mathbf{I}_{\phi^i}\left(\mathbf{z^i};\mathbf{s^i}\right)$$

$$=D_{KL}\left(q_{\phi^i}(h^i)||q_{\phi^i}(z^i)q_{\phi^i}(s^i)\right). \qquad (12)$$

In practice, the products of $q_{\phi^i}(z^i)$ and $q_{\phi^i}(s^i)$ are not analytical in nature. We introduce discriminator $\hat{D}(.)$ (see Figure 2) to distinguish samples from the joint distribution and the product of the marginal distribution, that is,

$$L_{ISM}\approx L^{e}_{IEM}=E_{q_{\phi^i}(h^i)}[log\frac{D(.)}{1-D(.)}]. \qquad (13)$$

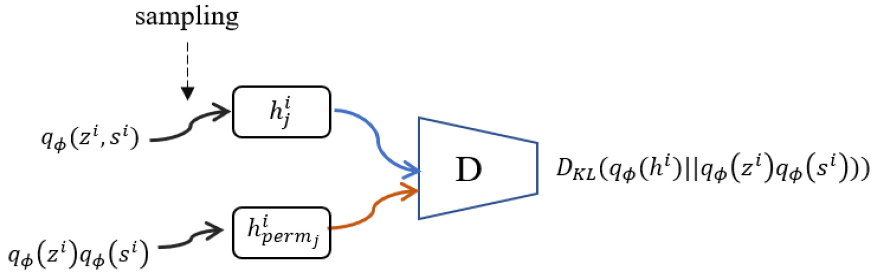


FIGURE 2
Discriminator *D*(.). To compute and optimize L_{ISM}, we need an additional discriminator as shown in Eq. (13).

FIGURE 3
Movements in NinaPro DB2. **(A)** Isometric, isotomic hand configurations. **(B)** Basic movements of the wrist. **(C)** Grasps and functional movements. **(D)** Single and multiple fingers force measurement patterns. **(E)** Rest position. Available from: http://ninapro.hevs.ch/node/123.

TABLE 1 Subject attribute information of NinaPro DB2 dataset.

Subject	Hand	Laterality	Gender	Age	Height (cm)	Weight (kg)
1	Intact	Right Handed	Male	29	187	75
2	Intact	Right Handed	Male	29	183	75
3	Intact	Right Handed	Male	31	174	69
4	Intact	Left Handed	Female	30	154	50
5	Intact	Right Handed	Male	25	175	70

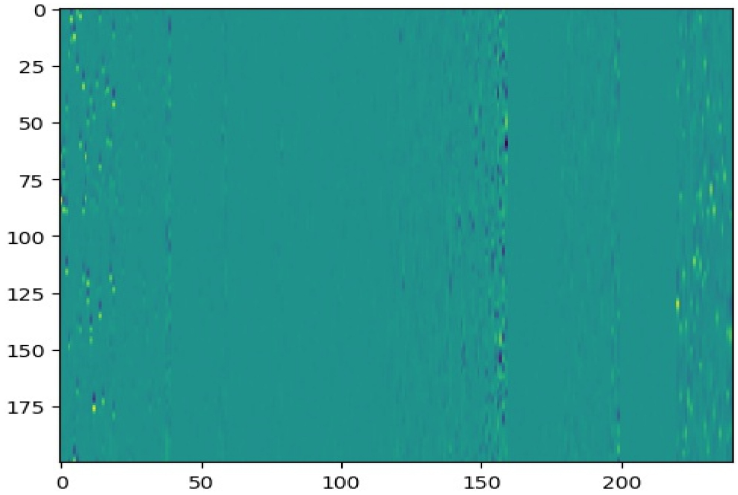


FIGURE 4
Sample data image.

TABLE 2 Detail parameters for LFEA.

Parameter	Value
Number of layers	4
Size of z^i	5
λ	0.1
β	0.2

TABLE 3 Results of TC score.

Method	TC score	MIG
LFEA (Ours)	**12.3**	**0.72**
VAE	23.6	0.54
β-VAE	25.8	0.61
PCA	18.5	0.49

We compare our method the classic methods including VAE, β-VAE and PCA. Our HFEA method is much better than others. The bold indicates the best results.

Algorithm optimization

As presented above, our model contains three modules: ICM, IEM, and ISM. However, during optimization, the back-propagation algorithm is computationally intensive and potentially problematic when training deep networks, so we propose a layer-wise training step. After training one layer of the network, we fix the parameters of the trained layers and only train the next layer in the next step. Finally, we can obtain the final model after training all the layers. Such optimization design allows for training parameters at the bottom layers without bac-propagation from the top layers, avoiding the problems that often occur with deep network optimization, like vanishing and exploding gradient.

Numerical results

Dataset

In our experiments, we used the NinaPro* DB2 dataset and DB5 dataset. Atzori et al. (2014), Gijsberts et al. (2014) as the benchmark to perform numerical comparisons. NinaPro is a standard dataset for the gesture recognition of sparse multichannel SEMG signals. The SEMG signals in DB2 were obtained from 40 subjects and included 49 types of hand movements (see **Figure 3**).

Detailed attribute information of the five subjects in NinaPro DB2 is shown in **Table 1**. The original SEMG signal was processed through sliding windows, and the size of the sample data used in the experiment was (200,12). **Figure 4** shows 20 processed data points.

DB1 consists of 11 subjects and the data set of each subject contains three types of gestures, which are Exercise A, Exercise B, and Exercise C. Exercise A includes 12 basic movements of fingers (see **Figure 5**). Exercise B includes 17 movements. Exercise C includes 23 grasping and functional movements.

We preprocessed the dataset with the digital filter to cutoff frequency and sliding window to split signal, which follows He et al. (2018).

Model setting

In the following experiments, we used four layers model. The loss function is as follows:

$$\min L \triangleq \mathbf{L}_{ICM}^{upper} - \lambda \mathbf{L}_{IEM}^{lower} + \beta \mathbf{L}_{ISM},$$

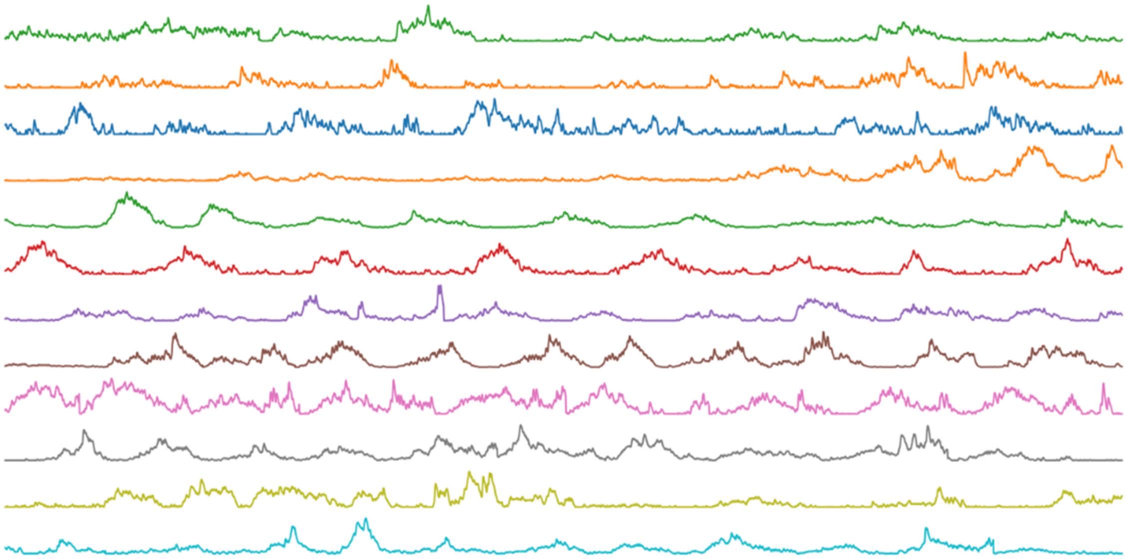


FIGURE 5
12 basic movements signal of fingers in Exercise A.

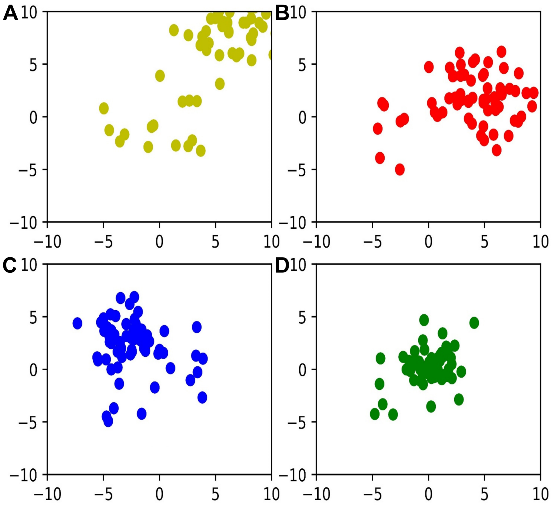


FIGURE 6
Feature distribution in layer 1–4 with **(A–D)**.

Detail parameters are listed in Table 2.

Results

First, we used total correlation (TC) as the quantitative metric for the quality of the disentanglement of the representation. TC is defined as follows:

$$TC\left(z^1, z^2, z^3, z^4\right) = E_{p(z^1,z^2,z^3,z^4)}\left[log\frac{p\left(z^1, z^2, z^3, z^4\right)}{p\left(z^1\right)p\left(z^2\right)p\left(z^3\right)p\left(z^4\right)}\right].$$

The TC was estimated using a three-like algorithm (Cheng et al., 2020a). A low TC score indicated that the representation had less variance. MIG metric (Chen et al., 2018) is another disentanglement metric; the higher the value, the more disentangled representation is. We compared the quality of disentanglement among PCA, β-VAE, VAE, and HFEA. Table 3 shows the comparison results on TC score and MIG

TABLE 4 Classification results on NinaPro DB2 dataset.

Methods	Windowing	Train/Test	Accuracy
LFEA + SVM(Ours)	200 ms	2/1	75.2 ± 2.3%
CNN	200 ms	2/1	65.7 ± 5.9%
LSTM + MLP	200 ms	1/1	**75.4 ± 8.2%**
Random forest	200 ms	2/1	75.0 ± 5.1%
KNN	200 ms	2/1	61.1 ± 3.4%
SVM	200 ms	2/1	67.2 ± 5.2%

The bold indicates better result.

metric. In TC score and MIG metric, HFEA has the best performance, which is 6.2 lower and 0.11 higher than the second place, respectively.

Furthermore, in Figure 6, we visualize the distribution of z^1, z^2, z^3, and z^4, respectively in a two-dimensional space based on t-distributed stochastic neighbor embedding. We can find that the variance of representation decreases with deeper layers, which indicates that the deeper networks learn more robust representations.

Classification results on NinaPro DB2 dataset is described in Table 4. Our method is based on LFEA and SVM and the feature Z used in SVM is computed by LFEA.

$$Z = \left(z^1, z^2, z^3, z^4\right)$$

The methods used for comparison include LSTM + CNN (He et al., 2018), k-nearest neighbor (KNN), support vector machine (SVM), random forest, and convolutional neural network (CNN) (Atzori et al., 2016). In all experiments, our method was second best in all methods and only 0.2% lower than the best. What is more, our method showed more stable results (2.3% fluctuations) than others.

Discrimination results for Exercise A, Exercise B, and Exercise C in DB1 and DB2 is shown in Figures 7, 8, respectively. For each exercise, we compare feature combinations from layer 1–4. Detail feature combinations is described in Table 5. Tables 6–8 list the classification accuracy with different feature combinations for DB1, respectively.

Discrimination value in Tables 6–8 measures the representation capability of feature in each layer. The higher the value, the better the feature representation ability. In Exercise A, C4 obtains the highest discrimination value, which means feature z^3 plays the most import role in Exercise A. Similarly, feature z^2 makes little difference in Exercise A.

Conclusion

In this manuscript, we propose an Unsupervised Layer-wise Feature Extraction Algorithm (LFEA) to perform the sEMG signal processing and downstream classification tasks. The model contains three core modules: Information Compression Module (ICM), Information Expression

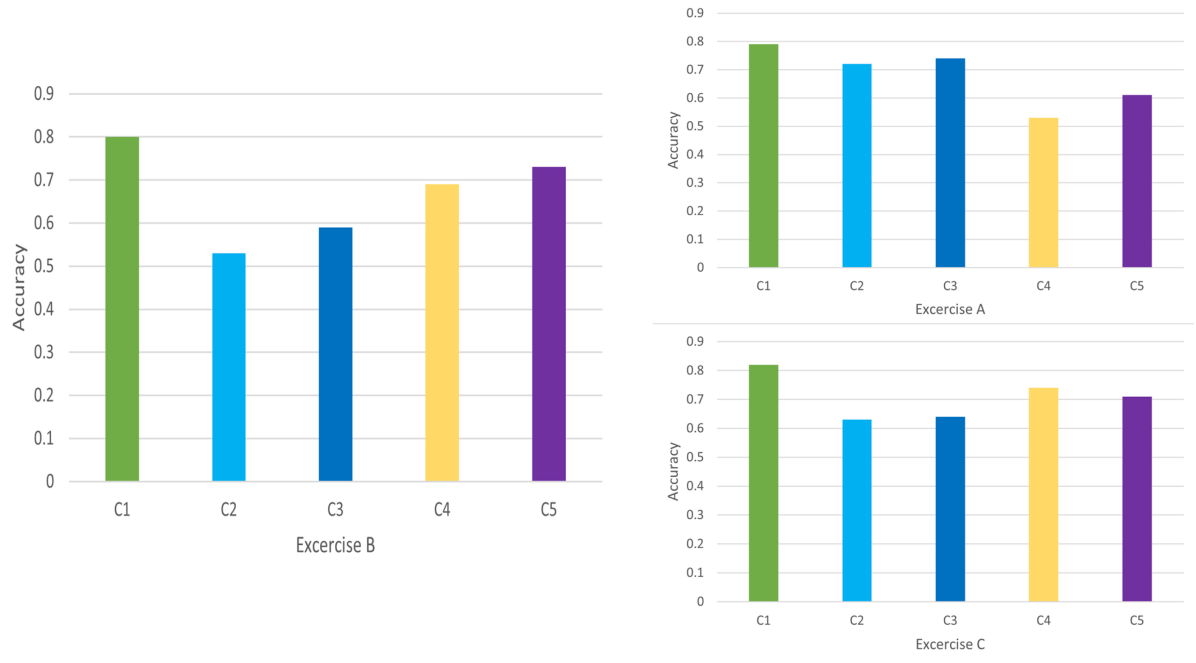


FIGURE 7
Feature discrimination results for DB1.

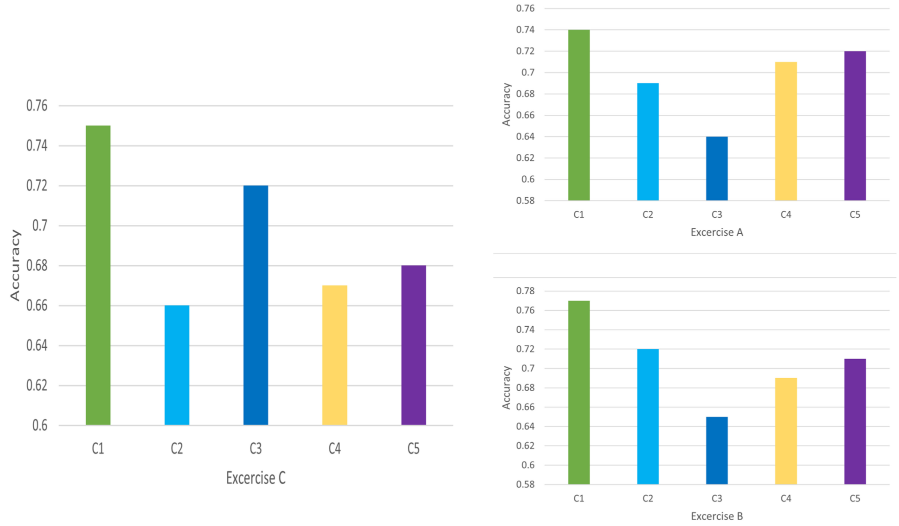


FIGURE 8
Feature discrimination results for DB2.

TABLE 5 Feature combinations.

C1	(z^1, z^2, z^3, z^4)
C2	(z^2, z^3, z^4)
C3	(z^1, z^3, z^4)
C4	(z^1, z^2, z^4)
C5	(z^1, z^2, z^3)

TABLE 6 Classification results with different feature combinations for Exercise A.

Feature Combinations	Accuracy	Discrimination (C1-Accuracy)
C1	0.79	0
C2	0.72	0.07
C3	0.74	**0.05**
C4	0.53	**0.26**
C5	0.61	0.18

The bold values mean the lowest and highest discrimination values.

TABLE 7 Classification results with different feature combinations for Exercise B.

Feature Combinations	Accuracy	Discrimination (-C1)
C1	0.8	0
C2	0.53	**0.27**
C3	0.59	0.21
C4	0.69	0.11
C5	0.73	**0.07**

The bold values mean the lowest and highest discrimination values.

TABLE 8 Classification results with different feature combinations for Exercise C.

Feature Combinations	Accuracy	Discrimination (-C1)
C1	0.82	0
C2	0.63	**0.19**
C3	0.64	0.18
C4	0.74	**0.08**
C5	0.71	0.11

The bold values mean the lowest and highest discrimination values.

Module (IEM) and Information Separation Module (ISM), that ensure that the learning representation is compact, informative and disentangled. We further use a layer-wise optimization procedure to reduce the computation cost and avoid some optimization problem, like vanishing and exploding gradient. Experimentally, we also verify that the untangling effect and downstream classification tasks give better results.

In the future, we hope to combine the advantages of supervised and unsupervised to build a semi-supervised learning framework that can be adapted to more scenarios.

Data availability statement

The original contributions presented in this study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

ML and ZL contributed to the conception and design of the study. FZ organized the database. JG performed the statistical analysis. ML and ST wrote the first draft of the manuscript. All authors contributed to the manuscript revision, read, and approved the submitted version.

Funding

This study was supported by the National Key R&D Program of China (2021YFA1000401) and the National Natural Science Foundation of China (U19B2040).

Conflict of interest

ML, ST, JG, and FZ were employed by Information Science Academy, China Electronics Technology Group Corporation.

The remaining 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.

Publisher's note

References

Abdi, H., and Williams, L. J. (2010). Principal component analysis. *Wiley Interdiscip. Rev.* 2, 433–459. doi: 10.1002/wics.101

Alemi, A. A., Fischer, I., Dillon, J. V., and Murphy, K. (2016). Deep variational information bottleneck. *arXiv*. [preprint]. arXiv:1612.00410.

Atzori, M., Cognolato, M., and Müller, H. (2016). Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands. *Front. Neurorobot.* 10:9. doi: 10.3389/fnbot.2016.00009

Atzori, M., Gijsberts, A., Castellini, C., Caputo, B., Hager, A. G. M., Elsig, S., et al. (2014). Electromyography data for non-invasive naturally-controlled robotic hand prostheses. *Sci. Data* 1:140053. doi: 10.1038/sdata.2014.53

Belghazi, M. I., Baratin, A., Rajeshwar, S., Ozair, S., Bengio, Y., Courville, A., et al. (2018). "Mutual information neural estimation," in *International Conference on Machine Learning*, 531–540. (Stockholmsmässan, Stockholm Sweden)

Belkin, M., and Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* 15, 1373–1396. doi: 10.1162/089976603321780317

Bell, A. J., and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1129–1159. doi: 10.1162/neco.1995.7.6.1129

Bengio, Y., Courville, A., and Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 1798–1828. doi: 10.1109/TPAMI.2013.50

Burgess, C. P., Higgins, I., Pal, A., Matthey, L., and Lerchner, A. (2018). Understanding disentangling ß in vae. *arXiv* [Preprint]. arXiv:1804.03599.

Chen, R. T., Li, X., Grosse, R. B., and Duvenaud, D. K. (2018). Isolating sources of disentanglement in variational autoencoders. *Adv. Neural Inf. Proc. Syst.* [Preprint]. arXiv:1802.04942.

Cheng, P., Min, M. R., Shen, D., Malon, C., Zhang, Y., Li, Y., et al. (2020b). Improving disentangled text representation learning with information-theoretic guidance. *arXiv* [preprint]. arXiv:2006.00693. doi: 10.18653/v1/2020.acl-main.673

Cheng, P., Hao, W., and Carin, L. (2020a). Estimating Total Correlation with Mutual Information Bounds. *arXiv* [Preprint]. arXiv:2011.04794.

Gijsberts, A., Atzori, M., Castellini, C., Müller, H., and Caputo, B. (2014). Measuring movement classification performance with the movement error rate. *IEEE Trans. Neural Syst. Rehabil. Eng.* 89621, 735–744. doi: 10.1109/TNSRE.2014.2303394

Gonzalez-Garcia, A., Van De Weijer, J., and Bengio, Y. (2018). Image-to-image translation for cross-domain disentanglement. *Adv. Neural Inf. Proc. Syst.* 31, 1287–1298

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (Berlin: Springer), 2672-2680.

Hassani, K., and Khasahmadi, A. H. (2020). Contrastive multi-view representation learing on graphs. *arXiv*. [Preprint]. arXiv:2006.05582.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (Washington, DC: IEEE Computer Society), 770–778. doi: 10.1109/CVPR.2016.90

He, Y., Fukuda, O., Bu, N., Okumura, H., and Yamaguchi, N. (2018). "Surface emg pattern recognition using long short-term memory combined with multilayer perceptron," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (Jeju Island: IEEE), 5636–5639. doi: 10.1109/EMBC.2018.8513595

Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., et al. (2016). "Beta-VAE: Learning basic visual concepts with a constrained variational framework," in *Proceedings of the international conference on learning representations*.

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

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., et al. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv*. [Preprint]. arXiv:1704.04861.

Hsu, W. N., Zhang, Y., and Glass, J. (2017). Unsupervised learning of disentangled and interpretable representations from sequential data. *Adv. Neural Inf. Proc. Syst.* [Preprint]. arXiv:1709.07902.

Izenman, A. J. (2013). "Linear discriminant analysis," in *Modern multivariate statistical techniques*, (New York, NY: Springer), 237–280. doi: 10.1007/978-0-387-78189-1_8

Jeon, I., Lee, W., Pyeon, M., and Kim, G. (2021). "Ib-gan: Disengangled representation learning with information bottleneck generative adversarial networks," in *Proceedings of the AAAI Conference on Computer Vision and Pattern Recognition*, 7926–7934.

Kim, H., and Mnih, A. (2018). "Disentangling by factorising," in *International Conference on Machine Learning*, 2649–2658.

Kingma, D. P., and Welling, M. (2013). Auto-encoding variational bayes. *arXiv*. [Preprint]. arXiv:1312.6114.1.

LeCun, Y., Touresky, D., Hinton, G., and Sejnowski, T. (1988). " A theoretical framework for back-propagation," in *In Proceedings of the 1988 Connectionist Models Summer School*, Vol. 1, 21–28.

Liu, Z., Li, M., and Han, C. (2021). Blocked and Hierarchical Disentangled Representation From Information Theory Perspective. *arXiv*. [preprint]. arXiv:2101.08408.

Matsuda, Y., and Yamaguchi, K. (2003). "The InfoMin criterion: An information theoretic unifying objective function for topographic mappings," in *Artificial Neural Networks and Neural Information Processing—ICANN/ICONIP 2003*, (Berlin: Springer), 401–408. doi: 10.1007/3-540-44989-2_48

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *J. Math. Psychol.* 47, 90–100. doi: 10.1016/S0022-2496(02)00028-7

Richard, M. D., and Lippmann, R. P. (1991). Neural network classifiers estimate Bayesian a posteriori probabilities. *Neural Comput.* 3, 461–483. doi: 10.1162/neco.1991.3.4.461

Shannon, C. E. (2001). A mathematical theory of communication. *GetMobile* 5, 3–55. doi: 10.1145/584091.584093

Shwartz-Ziv, R., and Tishby, N. (2017). Opening the black box of deep neural networks via information. *arXiv*. [Preprint]. arXiv:1703.00810.

Tenenbaum, J. B., Silva, V. D., and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science* 290, 2319–2323. doi: 10.1126/science.290.5500.2319

Thomas, M. T. C. A. J., and Joy, A. T. (2006). *Elements of information theory*. Hoboken, NJ: Wiley-Interscience.

Tishby, N., and Zaslavsky, N. (2015). "Deep learning and the information bottleneck principle," in *2015 IEEE Information Theory Workshop (ITW)*, (Jeju Island: IEEE), 1–5. doi: 10.1109/ITW.2015.7133169

Tishby, N., Pereira, F. C., and Bialek, W. (2000). The information bottleneck method. *arXiv*. [Preprint]. physics/0004057.

Xing, K., Ding, Z., Jiang, S., Ma, X., Yang, K., Yang, C., et al. (2018). "Hand gesture recognition based on deep learning method," in *2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)*, (Jeju Island: IEEE), 542–546. doi: 10.1109/DSC.2018.00087

Yingzhen, L., and Mandt, S. (2018). "Disentangled sequential autoencoder," in *International Conference on Machine Learning*, 5670–5679.

Zbontar, J., Jing, L., Misra, I., Lecun, Y., and Deny, S. (2021). Barlow twins: Self-supervised learning via redundancy reduction. *Int. Conference Mach. Learn.* 139, 12310–12320.

Zhai, X., Jelfs, B., Chan, R. H., and Tin, C. (2017). Self-recalibrating surface EMG pattern recognition for neuroprosthesis control based on convolutional neural network. *Front. Neurosci.* 11:379. doi: 10.3389/fnins.2017.00379

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 16 August 2022
DOI 10.3389/fnins.2022.973761

OPEN ACCESS

EDITED BY
Hui Zhou,
Nanjing University of Science and Technology, China

REVIEWED BY
Changming Wang,
Xuanwu Hospital, Capital Medical University, China
Mengchu Zhou,
New Jersey Institute of Technology, United States
Kazuyuki Matsumoto,
Tokushima University, Japan

*CORRESPONDENCE
Li Lei
leili@bupt.edu.cn

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

RECEIVED 20 June 2022
ACCEPTED 22 July 2022
PUBLISHED 16 August 2022

CITATION
Yubo Z, Yingying L, Bing Z, Lin Z and Lei L (2022) MMASleepNet: A multimodal attention network based on electrophysiological signals for automatic sleep staging.
Front. Neurosci. 16:973761.
doi: 10.3389/fnins.2022.973761

MMASleepNet: A multimodal attention network based on electrophysiological signals for automatic sleep staging

Zheng Yubo, Luo Yingying, Zou Bing, Zhang Lin and Li Lei*

School of Artificial Intelligence, University of Posts and Telecommunications, Beijing, China

Pandemic-related sleep disorders affect human physical and mental health. The artificial intelligence (AI) based sleep staging with multimodal electrophysiological signals help people diagnose and treat sleep disorders. However, the existing AI-based methods could not capture more discriminative modalities and adaptively correlate these multimodal features. This paper introduces a multimodal attention network (MMASleepNet) to efficiently extract, perceive and fuse multimodal features of electrophysiological signals. The MMASleepNet has a multi-branch feature extraction (MBFE) module followed by an attention-based feature fusing (AFF) module. In the MBFE module, branches are designed to extract multimodal signals' temporal and spectral features. Each branch has two-stream convolutional networks with a unique kernel to perceive features of different time scales. The AFF module contains a modal-wise squeeze and excitation (SE) block to adjust the weights of modalities with more discriminative features and a Transformer encoder (TE) to generate attention matrices and extract the inter-dependencies among multimodal features. Our MMASleepNet outperforms state-of-the-art models in terms of different evaluation matrices on the datasets of Sleep-EDF and ISRUC-Sleep. The implementation code is available at: https://github.com/buptantEEG/MMASleepNet/.

KEYWORDS

multimodal, attention network, automatic sleep staging, electrophysiological signals, features fusion

1. Introduction

Sleep is an essential natural behavior for humans to maintain mental and physical health. Surveys show that ordinary people worldwide also have insomnia attributed to pandemic-related stress, anxiety, depression, and other mental health conditions during the new coronavirus pandemic (Semyachkina-Glushkovskaya et al., 2021). Survivors of COVID-19 are still bothered by insomnia (Taquet et al., 2021). The research found that adequate and effective sleep helps people improve the efficacy of COVID-19 vaccines (Benedict and Cedernaes, 2021), and sleeping in the rapid eye movement (REM) stage helps restore the brain's ability and remove waste from the brain (Van Alphen et al., 2021). Sleep staging helps ordinary people better understand their sleep quality and helps patients with insomnia or other related diseases to obtain better diagnoses and treatment (Pan et al., 2020).

Polysomnography (PSG) is the primary tool for assessing sleep in the laboratory and can be used for clinical and research purposes (Rundo and Downey, 2019). During polysomnography, EEG, EOG, EMG, and other electrophysiological signals are recorded as multimodal data and then used by professional doctors to divide sleep into distinct stages. The American Academy of Sleep Medicine (AASM) classifies each 30 s sleep epoch into five different stages (W, N1, N2, N3, and REM) (Chriskos et al., 2021). However, manual sleep staging requires professional knowledge and is highly time-consuming. Artificial intelligence technology helps to improve efficiency and has become a research hot spot of sleep staging in recent years.

There have been two main approaches widely adopted in sleep staging studies. Some researchers employed conventional machine learning methods, which mainly contained feature extraction algorithms and fed features into conventional classifiers (Awais et al., 2021). Due to the need for prior professional knowledge for feature extraction, these models have poor transfer ability, and non-end-to-end learning is significantly subject to subjective influence. For other researchers, deep learning methods were adopted due to their superior performance and less need for prior knowledge. Some studies designed convolutional neural networks (CNNs) for sleep staging (Supratak et al., 2017; Phan et al., 2018; Perslev et al., 2019; Jia et al., 2020). Some studies employed long short-term memory (LSTM) to capture the temporal context from the representative features in forward and backward directions (Supratak et al., 2017; Supratak and Guo, 2020; Neng et al., 2021). Recurrent Neural Networks (RNNS) were proposed to capture the temporal correlation of electrophysiological signals (Michielli et al., 2019). Attention mechanism and attention-based feature fusion have been widely used in multimodal representation learning (Huang et al., 2019, 2020; Lu et al., 2019; Wei et al., 2020; Zhang et al., 2020a,b,c; Desai and Johnson, 2021; Yu et al., 2021; Ma et al., 2022). The existing studies based on attention mechanisms usually used single-modal data such as EEG or EOG, which only focused on the inter-relationship among single modality features rather than cross-modal features (Eldele et al., 2021).

The waveforms of EEG, EOG, and EMG in each sleep stage are shown in Figure 1. The signal characteristics of each modality among the five sleep stages are different, whether in the time domain or frequency domain. Observed from the time domain, signal amplitudes and cycles of different modalities signals are also various. Using EEG alone for sleep staging has been a feasible solution since EEG is the main basis of artificial sleep staging. It can also be observed that there are significant differences between the W stage and N1 stage in EOG waveforms, and the EMG waveforms are also helpful in identifying REM. Most studies chose EEG as the primary modality (Supratak et al., 2017). Some studies selected EOG signals which could be more convenient to acquire than EEG signals (Fan et al., 2021). Other studies also adopted EMG signals with more distinguishable features between the W and REM stages (Li et al., 2022). Further, it can be verified that the electrophysiological signals of the three modalities have complementary characteristics to sleep staging. By designing a neural network method of modality fusion, the accuracy of sleep staging can be improved. The existing multimodal sleep staging methods usually took EEG and EOG as the input of the model, and the fusion of multimodal features was mainly based on concatenation (Jia et al., 2020, 2021) without focusing on parts of the features.

To efficiently extract multimodal features of EEG, EOG, and EMG, use the attention mechanism for feature fusion, and improve the accuracy of sleep staging, the multimodal attention network (MMASleepNet) is proposed, which has a multi-branch feature extraction module followed by an attention fusing module, as shown in Figure 2. The contributions of this paper are as follows.

(1) The multi-branch feature extraction (MBFE) module is proposed, and unique kernels are specially designed based on the effective frequency band of three modalities.
(2) The attention-based Feature Fusion (AFF) module is proposed, and modal-wise squeeze and excitation block are combined with Transformer Encoder to fuse the features of EEG, EOG, and EMG.
(3) Experiments on four public datasets validate the effectiveness of the MMASleepNet. The results demonstrate that MMASleepNet outperforms all the baseline models in automatic sleep staging.

The context of this paper is as follows. Section 2 introduces data and methodology. The experiment design is described in the Section 3. Section 4 presents the results of experiments, and Section 5 analyses the results.

2. Materials and methods

2.1. Data description

Publicly available datasets were used for method evaluation, whose summary is shown in Table 1.

2.1.1. Sleep-EDF

The Sleep-EDF dataset contains two sub-datasets, namely, Sleep-EDF-20 and Sleep-EDF-78 (Goldberger et al., 2000). The Sleep-EDF-20 dataset contains 42308 epochs in 39 sleep cassette files collected from 20 subjects aged 25–34. The Sleep-EDF-78 dataset contains 195479 epochs in 153 sleep cassette files of 78 subjects aged 25–101. Each subject of the Sleep-EDF database contains 2 day-night PSG recordings except subjects 13, 36,

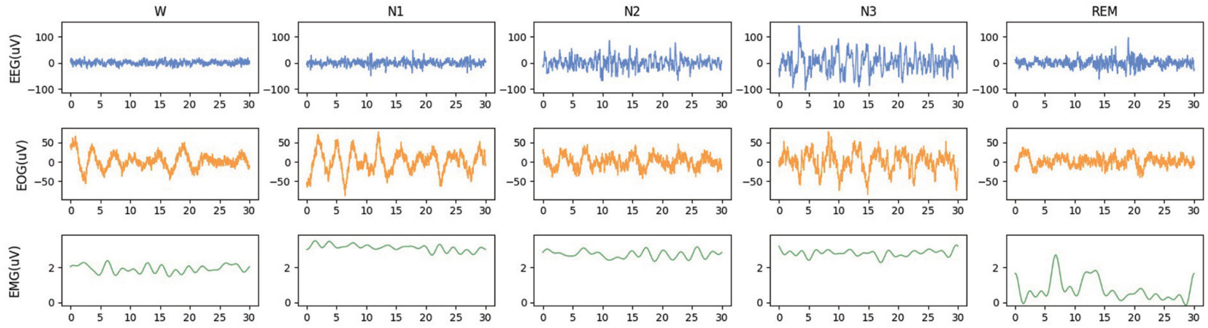


FIGURE 1
The waveforms of EEG, EOG, and EMG in each sleep stage. The data is randomly selected from the Sleep-EDF-78 dataset, and each epoch is 30 s.

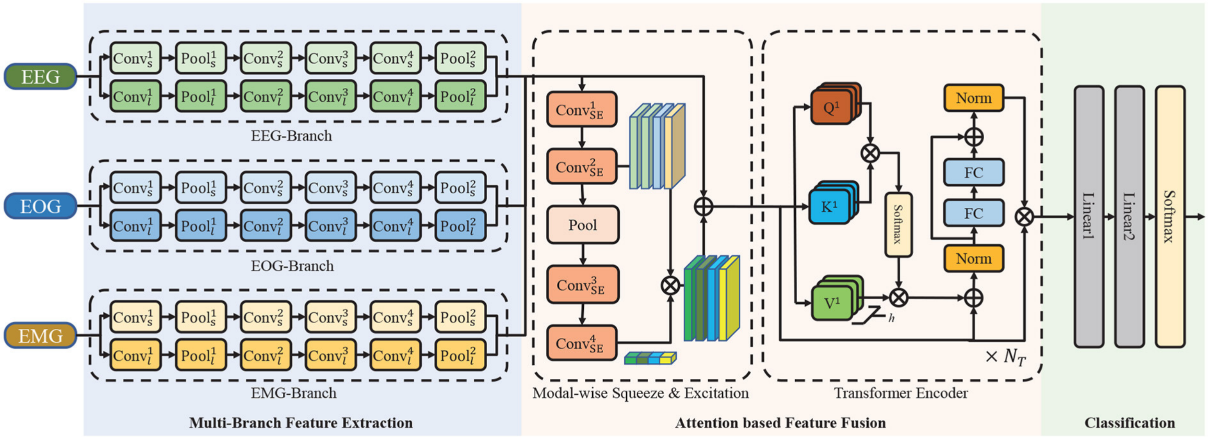


FIGURE 2
The architecture of the proposed network. It consists of a multi-branch feature extraction module, an attention based feature fusion module and a classification module. ⊕ is the point-wise addition and ⊗ is the point-wise multiplication. Conv is the convolutional layer, Pool is the pooling layer, FC is the fully connection layer, Norm is the normalization layer.

TABLE 1 Summary of the datasets and selected channels.

Dataset	Subjects	Samples	W (%)	N1 (%)	N2 (%)	N3 (%)	REM (%)	Score method	k for k-fold
Sleep-EDF-20	20	42,308	19.58	6.63	42.07	13.48	18.24	R&K	20
Sleep-EDF-78	78	195,479	33.74	11.01	35.37	6.67	13.22	R&K	10
ISRUC-Sleep-1	100	87,187	22.95	12.85	31.51	19.45	13.23	AASM	5
ISRUC-Sleep-3	10	8,589	20.44	14.04	30.12	22.90	12.50	AASM	10

and 52, whose one recording is lost due to device failure. The duration of each epoch is 30 s, and it has been labeled as {Wake, REM, N1, N2, N3, N4, MOVEMENT, UNKNOWN } by experts according to the R&K standard.

2.1.2. ISRUC-sleep

ISRUC-Sleep-1 and ISRUC-Sleep-3 are the sub-datasets of the ISRUC-Sleep (Khalighi et al., 2016). The ISRUC-Sleep-1 dataset contains 69,671 epochs in 100 PSG data files collected from 100 subjects aged 20–85. The ISRUC-Sleep-3 dataset contains 8,589 epochs in 10 PSG data files collected from 10 subjects aged 30–58. Each recording contains 6 EEG channels (F3-A2, C3-A2, O1-A2, F4-A1, C4-A1, and O2-A1), 2 EOG channels (LOC-A2 and ROC-A1), 3 EMG channels (Chin EMG, left leg movements and right leg movements), and 1 ECG channel, and all signals were sampled at 200 Hz. The duration of each epoch is 30 s, and it has been labeled as {Wake, REM, N1, N2, N3 } by experts according to AASM standard.

For a fair comparison with baseline models, the following data preprocessing steps have been applied to the Sleep-EDF and ISRUC-Sleep datasets. The N3 and N4 are merged into N3 according to the AASM standard for the Sleep-EDF dataset. Then, MOVEMENT and UNKNOWN epochs are excluded. The signals of EEG (Fpz-Cz and Pz-Oz), EOG (ROC-LOC), and EMG (CHIN1-CHIN2) are adopted. For the ISRUC-Sleep dataset, the signals of EEG (F3-A2, C3-A2, O1-A2, F3-A1, C4-A1, O2-A1), EOG (ROC-A1), and EMG (CHin-EMG) are adopted. For the four datasets, 30 min of wake epochs before and after sleep epochs are maintained to focus more on the sleep stages. In this study, all these signals are resampled at 100 Hz.

2.2. Method

Figure 2 illustrates the overall framework of MMASleepNet. The MMASleepNet consists of three main modules: multi-branch feature extraction (MBFE), attention-based feature fusion (AFF), and classification. The network can be trained and optimized using multimodal electrophysiological signals. Firstly, raw signals of each modality are processed into high-level features by the specially designed branches in the MBFE module. This module has several two-stream convolutional networks, which consist of a small kernel fully convolutional network (FCN) and a large kernel FCN to perceive features of different time scales. The AFF module includes a modal-wise squeeze and excitation (SE) block to adjust the weights of modalities with more discriminative features and TE layers to generate attention matrices and extract the inter-dependencies among multimodal features. Finally, the staging results can be obtained through the classification layer.

2.2.1. Multi-branch feature extraction

In order to extract the features from the original multimodal data (EEG, EOG, and EMG), two-stream convolutional network branches are designed in the MBFE module. Each branch in the MBFE module consists of two FCN streams with four convolutional layers and two Max-Pooling layers. Referring to previous studies, the different sizes of convolutional kernels capture different scale features, making the feature matrix more comprehensive (Supratak et al., 2017). One FCN stream adopts a large kernel, and the other adopts a small kernel at the first convolutional layer. As the electrophysiological signals are sampled at 100 Hz, the convolutional layer with a kernel size of 500 extracts low-frequency information using 5-s windows. On the contrary, the small convolutional layer with a kernel size of 50 extracts the high-frequency information and detailed features using half-second windows. As the modalities have different interesting frequency ranges, the size of the convolutional kernel in the EEG branch is twice that of EOG and EMG. Due to EEG having higher classification accuracy in most cases, the

TABLE 2 Parameters of the MBFE module. *Size*$_k$ is the size of convolutional kernel, *N* is the numbers of filters and d_M is the number of kernels at the last convolutional layer.

Branch	**Stream**	**Layers**						d_M
		Conv1D-1	**MaxPooling-1**	**Conv1D-2**	**Conv1D-3**	**Conv1d-4**	**MaxPooling-2**	
EEG	Small	$Size_k = 50$ Stride = 8 $N = 64$	$Size_k = 8$ Stride = 8	$Size_k = 8$ Stride = 1 $N = 128$	$Size_k = 8$ Stride = 1 $N = 128$	$Size_k = 8$ Stride = 1 $N = 128$	$Size_k = 4$ Stride = 4	128
	Large	$Size_k = 500$ Stride = 64 $N = 64$	$Size_k = 4$ Stride = 4	$Size_k = 6$ Stride = 1 $N = 128$	$Size_k = 6$ Stride = 1 $N = 128$	$Size_k = 6$ Stride = 1 $N = 128$	$Size_k = 2$ Stride = 2	128
EOG	Small	$Size_k = 25$ Stride = 8 $N = 32$	$Size_k = 8$ Stride = 8	$Size_k = 8$ Stride = 1 $N = 64$	$Size_k = 8$ Stride = 1 $N = 64$	$Size_k = 8$ Stride = 1 $N = 64$	$Size_k = 4$ Stride = 4	64
	Large	$Size_k = 250$ Stride = 64 $N = 32$	$Size_k = 4$ Stride = 4	$Size_k = 6$ Stride = 1 $N = 64$	$Size_k = 6$ Stride = 1 $N = 64$	$Size_k = 6$ Stride = 1 $N = 64$	$Size_k = 2$ Stride = 2	64
EMG	Small	$Size_k = 25$ Stride = 8 $N = 32$	$Size_k = 8$ Stride = 8	$Size_k = 8$ Stride = 1 $N = 64$	$Size_k = 8$ Stride = 1 $N = 64$	$Size_k = 8$ Stride = 1 $N = 64$	$Size_k = 4$ Stride = 4	64
	Large	$Size_k = 250$ Stride = 64 $N = 32$	$Size_k = 4$ Stride = 4	$Size_k = 6$ Stride = 1 $N = 64$	$Size_k = 6$ Stride = 1 $N = 64$	$Size_k = 6$ Stride = 1 $N = 64$	$Size_k = 2$ Stride = 2	64

number of convolutional kernels d_{EEG} for the EEG branch is also larger than d_{EOG} and d_{EMG} for EOG and EMG branches. The parameters of the MBFE module are given in Table 2. The leaky rectified linear unit (Leaky-ReLU) is employed as the activation function of each convolutional layer, which can be defined as follows:

$$LeakyReLU(x) = \begin{cases} x, & x \geq 0 \\ \alpha x, & x < 0 \end{cases} \quad (1)$$

The Leaky-ReLU can solve the zero gradient vanishing problems for negative values, which are essential for the following modules. Dropout layers are applied after the first Max-Pooling in both streams and after the concatenation of both streams to reduce overfitting. The input $X_M \in \mathbb{R}^{3\times C_M \times N}$ are fed into the MBFE module for extracting the multimodal features where $M \in \{EEG, EOG, EMG\}$ represent modalities, and C_M is the number of channels for modal M, $N = SampleRate \times EpochTime$ is the samples for a single channel in one epoch. The operation is formalized as follows:

$$F_M = FCN^s_M(X_M) \| FCN^l_M(X_M) \quad \in \mathbb{R}^{d_M \times l} \quad (2)$$

where FCN^s and FCN^l represent FCN stream with the small and large convolutional kernel and $\|$ is the concatenate operation. $F_{EEG} \in \mathbb{R}^{d_{EEG}\times l}$ is divided into $F^i_{EEG} \in \mathbb{R}^{d\times l}$, where $d = \frac{1}{2}d_{EEG}, i \in \{1, 2\}$ to align with the F_{EOG} and F_{EMG} for the following concatenate operation. A new dimension is created on each modalities' features, and the concatenate operation is formalized as follows:

$$F = [F^1_{EEG}, F^2_{EEG}, F_{EOG}, F_{EMG}] \quad \in \mathbb{R}^{4\times d\times l} \quad (3)$$

where $[\cdot]$ is the concatenate operation on the newly created modal dimension. A feature map F that contains different modalities of information is obtained through above operations.

2.2.2. Attention-based feature fusion

The AFF module is designed for fusing features extracted by the MBFE module. The architecture shown in Figure 2 is designed based on attention methods. AFF module consists of a modal-wise SE block and TE layers.

2.2.2.1. Modal-wise SE

The modal-wise SE block is proposed based on the SENet (Hu et al., 2020). Different from the SENet using 1D convolutional and Max-Pooling layers, as shown in Figure 2, 2D convolutional and Max-Pooling layers are implemented to reconstruct the input features. Given a feature map $F \in \mathbb{R}^{4\times d\times l}$, two convolution operations are applied to F such that $F' = Conv_2(Conv_1(F))$ and F' has the exact dimensions as the input feature map. Global Average Pooling is performed along the spatial dimensions, and F' is turned into $S = \{S_1, S_2, S_3, S_4\}$. Two additional 2D convolutional layers replace the full connection layers in SENet to reconstruct S further. The first layer followed with ReLU activation function designed to reduce the dimensions of F, and the second layer followed with Sigmoid layer aims to increase the dimensions. The operation is formalized as follows:

$$E = \text{Sigmoid}(\text{Conv}_2(\text{ReLU}(\text{Conv}_1(S)))) \quad \in \mathbb{R}^{4\times d\times l} \quad (4)$$

where $Conv_1$ and $Conv_2$ are the 2D convolution operations, sigmoid and ReLU are the activation functions and $\text{ReLU}(x) = \max(0, x)$. The output dimension matches the number of input modalities. It characterizes the global distribution of responses over features. Then, the feature map F is scaled by E:

$$O_{SE} = F \oplus (F \otimes E) \quad \in \mathbb{R}^{4\times d\times l} \quad (5)$$

where $\oplus$ is the point-wise addition and $\otimes$ is the point-wise multiplication, O_{SE} is the output of the modal-wise SE block. Modal-wise SE block adaptively learns the correlation among multiple modalities and the attention of different modalities.

2.2.2.2. Transformer encoder

As shown in Figure 1, each TE layer comprises two core modules: multi-head attention and position-wise feed-forward network. Multi-head attention consists of H attention modules. Firstly, H different linear projections are applied to the input, and the result is mapped to parallel queries, keys, and values. Secondly, dot-product is performed on Q_i and K_i to calculate a similarity score. A normalization operation is applied to stabilize the gradient. Then, the Softmax operation calculates the weight for V_i, and another dot-product is applied. Finally, all the A_i are concatenated together to produce the final output. The operations can be formulated as follows:

$$Q_i = ZW^Q_i, K_i = ZW^K_i, V_i = ZW^V_i, 0 < i \leq H \quad (6)$$

$$A_i = \text{Softmax}(\frac{Q_i \cdot K^T_i}{\sqrt{d}}) \cdot V_i \quad (7)$$

$$MA = A_1 \| A_2 \| ... \| A_H \quad (8)$$

where $Z \in \mathbb{R}^{4l\times d}$ is the input of the TE layer. W^Q_i, W^K_i, $W^V_i \in \mathbb{R}^{d\times \frac{d}{H}}$ are learnable weights of linear projections, d is the column length of Z, and $\|$ is the concatenate operation. Residual layers are applied as Equation 9. The position-wise feed-forward network consists of two linear transformations with ReLU activation as follows:

$$O^1 = \text{LayerNorm}(MA + Z) \quad (9)$$

$$O^2 = \text{ReLU}(O^1 W_1 + b_1)W_2 + b_2 \quad (10)$$

where $W_1 \in \mathbb{R}^{d\times d_{FF}}$, $W_2 \in \mathbb{R}^{d_{FF}\times d}$ are learnable weight matrices $b_1 \in \mathbb{R}^{d_{FF}}$, $b_2 \in \mathbb{R}^{d}$ is learnable biases. d_{FF} is the middle dimension of the feed-forward network. Then the output of the attention-based feature fusion module O_{AF} can be obtained as follow:

$$O_{TE} = \text{LayerNorm}(O^1 \oplus O^2) \tag{11}$$

$$O_{AF} = Z \otimes O_{TE} \tag{12}$$

where Z is the flattened output of modal-wise SE block, and O_{TE} is the output of the TE layer. Then the O_{AF} is fed into two linear layers for the final classification.

3. Experiment

3.1. Baseline methods

Our method has been compared with the three baseline models: AttnSleepNet, SleepPrintNet, and SalientSleepNet. The publicly available codes have been used for AttnSleepNet, whereas SleepPrintNet and SalientSleepNet were re-implemented. For a fair comparison, all models were trained and tested on the same data partition with the same random seeds. Brief descriptions for models are as follows:

- **AttnSleepNet (Eldele et al., 2021):** AttnSleepNet deploys a custom CNN architecture followed by a multi-head attention mechanism and causal convolutions.
- **SleepPrintNet (Jia et al., 2020):** An EEG temporal feature extraction module, an EEG spectral-spatial feature extraction module, and two multimodal feature extraction modules are combined and classified.
- **SalientSleepNet (Jia et al., 2021):** A fully convolutional network based on the U^2-Net architecture. Two independent U^2-like streams are composed to extract the features from multimodal data.

3.2. Experiment settings

To evaluate the performance of models, subjects in each dataset were divided into several groups using k-fold cross-validation. For each fold, one group of subjects was selected as validation data. The remaining k-1 groups were selected as training data. Finally, four performance matrices were calculated by combining the predicted sleep stages of all k test groups. For the MMASleepNet, the Adam optimizer with the learning rate of 1e-4 was applied. The weight decay of Adam was set to 1e-3, the betas (b1, b2) were used as (0.9, 0.999), respectively, and the epsilon value was 1e-08. The parameters of the MBFE module are introduced in Table 1. The TE block has only one encoder layer with four heads. The training epoch is 150. Weighted cross-entropy loss was adopted as follows:

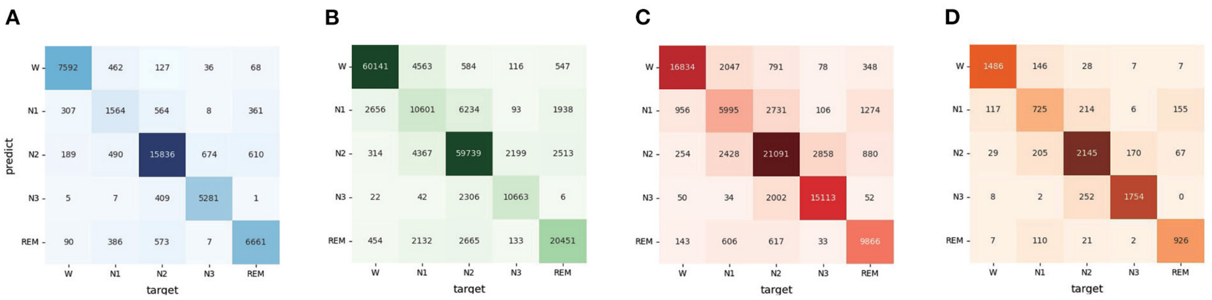
$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C}\omega_c y_i^c \log(p_i^c) \tag{13}$$

where N is the batch size, C is the number of classes, y_i^c is the true label, and p_i^c is the predicted label of i-th samples for class c. $\omega_c \in \{1.0, 1.80, 1.0, 1.25, 1.20\}$ is the weight of class c.

TABLE 3 Comparison among MMASleepNet and baseline models.

Dataset	Method	Per-class F1-score					Overall matrices			
		W	N1	N2	N3	REM	*ACC*	*MF1*	*κ*	*MGm*
Sleep-EDF-20	AttnSleepNet	79.02	32.70	87.03	85.67	72.36	79.10	71.35	71.43	66.34
	SleepPrintNet	88.77	47.99	86.72	86.21	80.26	83.08	77.99	76.67	76.34
	SalientSleepNet	90.79	49.86	89.03	84.77	**88.44**	86.28	80.58	81.02	77.32
	MMASleepNet	**92.20**	**54.75**	**89.70**	**90.20**	86.41	**87.30**	**82.65**	**82.63**	**81.67**
Sleep-EDF-78	AttnSleepNet	92.08	36.98	84.70	**81.63**	73.61	81.12	73.80	73.75	68.64
	SleepPtintNet	92.65	47.39	83.59	79.97	78.75	81.64	76.47	74.70	74.27
	SalientSleepNet	92.28	**50.52**	84.37	71.17	**84.19**	82.61	76.51	75.92	73.42
	MMASleepNet	**92.85**	49.05	**84.94**	81.26	79.75	**82.67**	**77.60**	**76.12**	**76.06**
ISRUC-SLEEP-1	AttnSleepNet	84.19	43.80	71.52	81.93	61.12	71.65	68.53	63.70	67.43
	SleepPtintNet	79.12	40.12	58.22	68.80	73.67	65.40	63.99	56.02	62.47
	SalientSleepNet	85.24	51.34	76.41	83.50	79.25	76.95	75.15	70.31	74.25
	MMASleepNet	**87.83**	**54.03**	**77.05**	**85.29**	**83.31**	**79.02**	**77.51**	**73.02**	**76.79**
ISRUC-SLEEP-3	AttnSleepNet	67.58	26.91	66.31	84.08	54.33	64.24	59.85	54.88	55.83
	SleepPrintNet	85.15	52.53	74.95	87.28	74.84	76.88	74.95	70.29	73.69
	SalientSleepNet	78.37	50.64	77.33	**87.99**	75.47	76.11	73.96	69.39	73.20
	MMASleepNet	**88.87**	**59.57**	**82.00**	87.00	**86.87**	**81.92**	**80.64**	**76.79**	**80.00**

The best values on each dataset are highlighted in bold.



A

predict \ target	W	N1	N2	N3	REM
W	7592	462	127	36	68
N1	307	1564	564	8	361
N2	189	490	15836	674	610
N3	5	7	409	5281	1
REM	90	386	573	7	6661

B

	W	N1	N2	N3	REM
W	60141	4563	584	116	547
N1	2656	10601	6234	93	1938
N2	314	4367	59739	2199	2513
N3	22	42	2306	10663	6
REM	454	2132	2665	133	20451

C

	W	N1	N2	N3	REM
W	16834	2047	791	78	348
N1	956	5995	2731	106	1274
N2	254	2428	21091	2858	880
N3	50	34	2002	15113	52
REM	143	606	617	33	9866

D

	W	N1	N2	N3	REM
W	1486	146	28	7	7
N1	117	725	214	6	155
N2	29	205	2145	170	67
N3	8	2	252	1754	0
REM	7	110	21	2	926

FIGURE 3
The confusion matrices of MMASleepNet, **(A)** is the confusion matrix valuated on SleepEDF-20 dataset, **(B)** is the confusion matrix valuated on SleepEDF-78 dataset, **(C)** is the confusion matrix valuated on ISRUC-Sleep-1 dataset, **(D)** is the confusion matrix valuated on ISRUC-Sleep-3 dataset.

For a fair comparison, all baseline models and proposed methods used the same dataset partitioning during training and evaluation. A number of experiments were conducted to find the best hyperparameters of the proposed MMASleepNet. The hyper-parameters of baseline models were set as introduced best in their article or open source codes. The train and validation codes are available at https://github.com/buptantEEG/MMASleepNet/.

3.3. Evaluation matrices

Four matrices were adopted to evaluate the performance of sleep staging models, namely, accuracy (ACC), macro-averaged F1-score (MF1), Cohen Kappa (κ), and the macro-averaged G-mean (MGm). Given True Positives (TP_i), False Positives (FP_i), True Negatives (TN_i), and False Negatives (FN_i) for the i-th class, the overall accuracy of ACC, MF1, κ, and MGm are defined as follows:

$$ACC = \frac{\sum_{c=1}^{C} TP_c}{N} \tag{14}$$

$$\kappa = \frac{ACC - p_e}{1 - p_e} \tag{15}$$

$$MF1 = \frac{1}{C} \sum_{c=1}^{C} \frac{2 \times Precision_c \times Recall_c}{Precision_c + Recall_c} \tag{16}$$

$$MGm = \frac{1}{C} \sum_{c=1}^{C} \sqrt{Specificity_c \times Recall_C} \tag{17}$$

where $p_e = \frac{\sum_{c=1}^{C} a_c \times b_c}{N \times N}$, $Precision_c = \frac{TP_c}{TP_c + FP_c}$, $Recall_c = \frac{TP_c}{TP_c + FN_c}$ and $Specificity_c = \frac{TN_c}{TN_c + FP_c}$, a_c is the number of samples of class c, b_c is the number of samples predicted as the class c. C is the number of classes, and N is the total number of samples.

4. Results

4.1. Results comparison with baselines

Table 3 shows the comparison among AttnSleepNet, SleepPrintNet, SalientSleepNet, and our MMASleepNet. The single-modal method AttnSleepNet obtained the lowest accuracy of the four models. The multimodal approaches, SleepPrintNet and SalientSleepNet, achieve higher accuracy than the single-modal method. The multimodal model can capture different electrophysiological signal features diversity compared to single-modal signals. In addition, the accuracy of the proposed MMASleepNet reaches 87.30, 82.67, 79.02, and 81.92%, which is higher than all the baseline models. The MF1, κ, and MGm of MMASleepNet outperform all baseline models on the four datasets, which means that the MMASleepNet is better at adapting to unbalanced data and should get better accuracy when the classes are balanced.

According to the confusion matrix in Figure 3, the classification accuracy of W, N2, N3, and REM is relatively high both on the Sleep-EDF dataset and ISRUC dataset. The accuracy of recognizing stage N1 is lower than in other stages, which is related to the insufficient N1 samples in the sleep records. Table 3 shows that the MMASleepNet obtained a higher F1 score for stage N1 on the smaller datasets Sleep-EDF-20, ISRUC-Sleep-1, and ISRUC-Sleep-3, indicating that the MMASleepNet performs better than the baseline methods for imbalanced categories. The results demonstrate the advantages of MMASleepNet in automatic sleep staging with the proposed feature extracting and fusion operations applied to multimodal electrophysiological signals.

4.2. Ablation experiments

The MMASleepNet consists of an MBFE module, modal-wise SE block, and Transformer Encoder layers. To analyze the

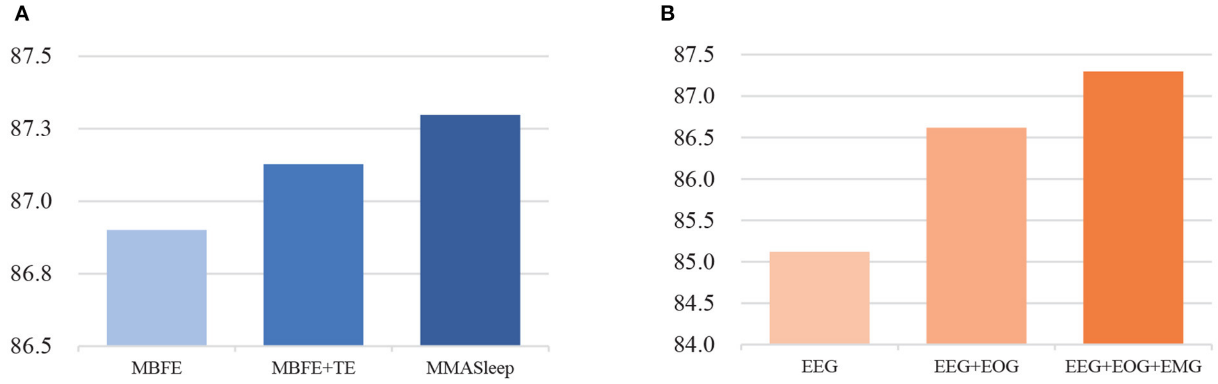


FIGURE 4
The results of ablation experiments, panel **(A)** is for the module ablation, panel **(B)** is for the modalities ablation.

influence of each module and to prove the effectiveness of each modality used in MMASleepNet, the ablation experiment was designed on the Sleep-EDF-20 dataset as follows:

- **MBFE(basic):** This model is only MBFE module input with EEG, EOG, and EMG signals. The features obtained from MBFE are fed into a linear classification module for sleep staging.
- **MBFE+TE:** This model adds TE layers based on the basic model input with EEG, EOG, and EMG signals.
- **MMASleepNet1:** The completely MMASleepNet with MBFE, modal-wise SE block, and TE layers, only input with EEG signals.
- **MMASleepNet2:** MMASleepNet input with EEG and EOG signals.
- **MMASleepNet3:** MMASleepNet input with EEG, EOG, and EMG signals.

Figure 4 presents the results of ablation experiments. Figure 4A shows that the attention-based feature fusion module improves the performance of the basic model. The modal-wise SE block helps the MMASleepNet achieve higher accuracy than only using TE layers. Figure 4B shows that MMASleepNet input with more modalities achieves higher accuracy. The model training with EOG and EMG performed better than with EEG alone.

5. Discussion

This study proposes a multimodal attention network for sleep staging using EEG, EOG, and EMG. The basis of using EEG, EOG, and EMG for sleep staging is that the PSG data collected in sleep health monitoring commonly includes multimodal electrophysiological signals. According to the experimental results, there are complementary features related to sleep stages among multiple modalities. The result shows that the proposed MMASleepNet achieves the highest classification performance on four publicly available datasets. Compared with the single-modality model AttnSleepNet, the proposed MMASleepNet can be fed with more data of multiple modalities, which means more information to extract and leads to big improvements in four evaluation matrices. Compared with the multimodal methods SleepPrintNet and SalinetSleepNet, MMASleepNet contains better-designed feature extraction methods and feature fusion methods for multimodal electrophysiological signals. The modal-wise SE block construct fusion of features adopted 2D convolutional, which makes it reasonable for complementary modalities. The SalientSleepNet also achieves high accuracy, but the high complexity of the modal led to lower training speed. The number of MMASleepNet parameters is 1.5M. The MMASleepNet has lower computation complexity and floating-point operations, improving the training speed. The AttnSleepNet, SleepPrintNet, SalientSleepNet, and the proposed MMASleepNet cost 0.4, 0.9, 7, and 1 h for 100 training epochs on the NVIDIA GeForce RTX 2080 Ti, respectively. Considering the accuracy and the training speed, the MMASleepNet performs better.

The ablation experiment results verify each module's effectiveness in the proposed MMASleepNet for automatic sleep staging. The ablation experiments in the first step verified that MMASleepNet fed with the data of three modalities achieves better results than a single modality. This preliminary verifies that the data of different modalities correlate with sleep stages and can be combined to obtain more time-frequency information. Features extracted from EOG and EMG complement those extracted from EEG only.

Figure 5 shows the down-sampled features before and after the AFF module. The main difference is that the features become more focused after the AFF module. The features after attention are easier to be distinguished using the same classifier, and the classifier is easier to converge. The visualized features show that the separability of the fused multimodal features can be

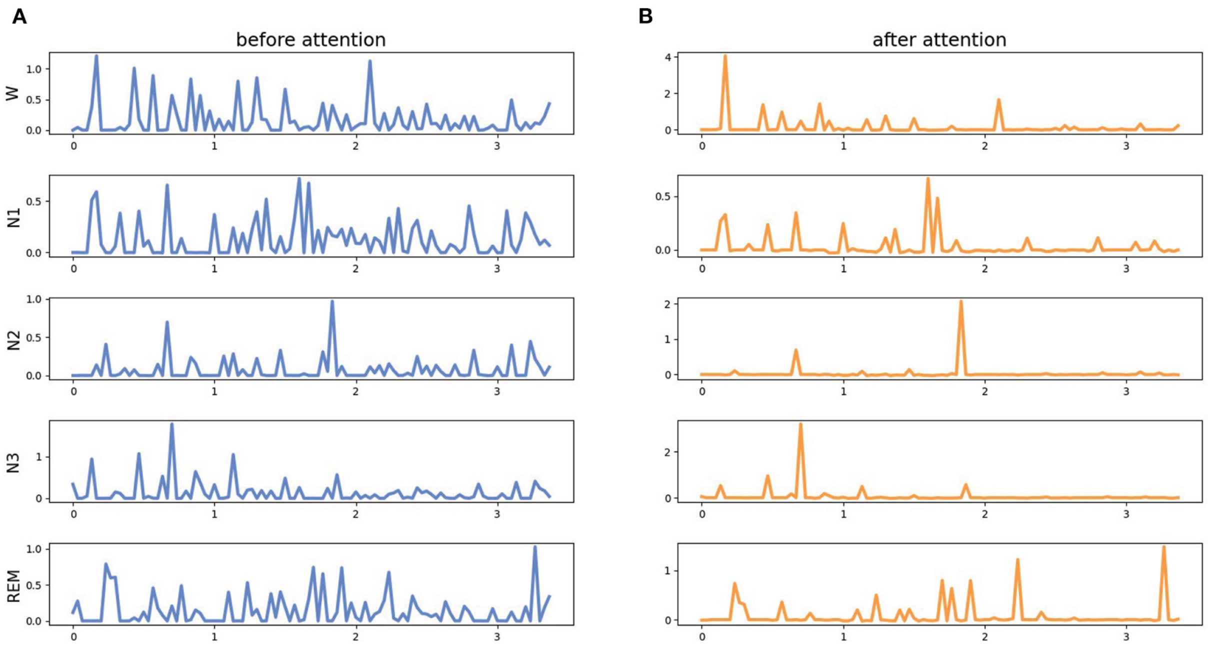


FIGURE 5
The features before and after attention mechanism of MMASleepNet. The data was selected randomly from the Sleep-EDF-20 dataset. Panel **(A)** is before the attention module, panel **(B)** is for the modalities ablation.

enhanced with the attention mechanism, and the neural network observes more detailed differences.

Unlike previous studies, the proposed MMASleepNet has a more effective feature fusion module, especially the modal-wise SE block, rather than a simple concatenate operation on different modalities' features. Although the model complexity has increased slightly, the model understands the relationship among different modalities. MMASleepNet can extract effective information from different modalities and fully use multimodal information by fusing the features with attention methods. Experiment results show that MMASleepNet achieves state-of-the-art performance. A series of ablation experiments have shown that different modules of the model contribute to the sleep staging task. MMASleepNet improves the accuracy of sleep staging, which provides a solution for multimodal sleep monitoring and is helpful for people to understand sleep status and improve their sleep quality.

Sleep disturbances increased significantly during the pandemic (Semyachkina-Glushkovskaya et al., 2021). For studying whether there has been a change in sleep disturbances, new sleep data set during the pandemic and new methods need to be supported, and deeper analysis based on statistical principles is required. Sleep staging is a fundamental application that helps study sleep disturbances during a real pandemic. With the existing standards, the definition of sleep stages will not be easily changed. Data set during the pandemic will be collected, and new methods for deeper analysis will be devised to raise awareness of the pandemic. The interpretability of the model and transfer learning method will be investigated to improve generalization across subjects and datasets.

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. Publicly available datasets were analyzed in this study. The Sleep-EDF can be found in https://www.physionet.org/content/sleep-edfx/1.0.0/ and the ISRUC-Sleep can be found in https://sleeptight.isr.uc.pt/.

Author contributions

ZY, LY, and ZB contributed to the conception of the study. ZY and LY realized proposed model and baseline models, performed the experiment, wrote the manuscript. ZY contributed significantly to analysis. LY and ZB contributed to the data curation, data analyses, and manuscript preparation. ZB helped to perform the analysis with constructive discussions. ZL and LL contributed to the funding acquisition and supervision. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 62176024) and project A02B01C01-201916D2.

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

References

Awais, M., Long, X., Yin, B., Abbasi, S. F., Akbarzadeh, S., Lu, C., et al. (2021). A hybrid DCNN-SVM model for classifying neonatal sleep and wake states based on facial expressions in video. *IEEE J. Biomed. Health Inform.* 25, 1441–1449. doi: 10.1109/JBHI.2021.3073632

Benedict, C., and Cedernaes, J. (2021). Could a good night's sleep improve COVID-19 vaccine efficacy? *Lancet Respirat. Med.* 9, 447–448. doi: 10.1016/S2213-2600(21)00126-0

Chriskos, P., Frantzidis, C. A., Nday, C. M., Gkivogkli, P. T., Bamidis, P. D., and Kourtidou-Papadeli, C. (2021). A review on current trends in automatic sleep staging through bio-signal recordings and future challenges. *Sleep Med. Rev.* 55, 101377. doi: 10.1016/j.smrv.2020.101377

Desai, K., and Johnson, J. (2021). "Virtex: learning visual representations from textual annotations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11162–11173. doi: 10.1109/CVPR46437.2021.01101

Eldele, E., Chen, Z., Liu, C., Wu, M., Kwoh, C.-K., Li, X., et al. (2021). An attention-based deep learning approach for sleep stage classification with single-channel EEG. *IEEE Trans. Neural Syst. Rehabil. Eng.* 29, 809–818. doi: 10.1109/TNSRE.2021.3076234

Fan, J., Sun, C., Long, M., Chen, C., and Chen, W. (2021). EOGNet: a novel deep learning model for sleep stage classification based on single-channel EOG signal. *Front. Neurosci.* 15, 573194. doi: 10.3389/fnins.2021.573194

Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., et al. (2000). PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation.* 101, e215–e220. doi: 10.1161/01.cir.101.23.e215

Hu, J., Shen, L., Albanie, S., Sun, G., and Wu, E. (2020). Squeeze-and-excitation networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 2011–2023. doi: 10.1109/TPAMI.2019.2913372

Huang, Z., Xu, X., Ni, J., Zhu, H., and Wang, C. (2019). Multimodal representation learning for recommendation in internet of things. *IEEE Internet Things J.* 6, 10675–10685. doi: 10.1109/JIOT.2019.2940709

Huang, Z., Xu, X., Zhu, H., and Zhou, M. (2020). An efficient group recommendation model with multiattention-based neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* 31, 4461–4474. doi: 10.1109/TNNLS.2019.2955567

Jia, Z., Lin, Y., Wang, J., Wang, X., Xie, P., and Zhang, Y. (2021). "SalientSleepNet: Multimodal salient wave detection network for sleep staging," in *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, ed Z. -H. Zhou (International Joint Conferences on Artificial Intelligence Organization), 2614–2620. doi: 10.24963/ijcai.2021/360

Jia, Z., Cai, X., Zheng, G., Wang, J., and Lin, Y. (2020). SleepPrintNet: a multivariate multimodal neural network based on physiological time-series for automatic sleep staging. *IEEE Trans. Artif. Intell.* 1, 248–257. doi: 10.1109/TAI.2021.3060350

Khalighi, S., Sousa, T., Santos, J. M., and Nunes, U. (2016). ISRUC-Sleep: A comprehensive public dataset for sleep researchers. *Comput. Methods Prog. Biomed.* 124, 180–192. doi: 10.1016/j.cmpb.2015.10.013

Li, Y., Xu, Z., Zhang, Y., Cao, Z., and Chen, H. (2022). Automatic sleep stage classification based on two-channel EOG and one-channel EMG. *Physiol. Meas.* 43. doi: 10.1088/1361-6579/ac6bdb

Lu, J., Batra, D., Parikh, D., and Lee, S. (2019). Vilbert: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks. *Advances in neural information processing systems* 32.

Ma, J., Tang, L., Fan, F., Huang, J., Mei, X., and Ma, Y. (2022). SwinFusion: cross-domain long-range learning for general image fusion via swin transformer. *IEEE/CAA J. Automat. Sin.* 9, 1200–1217. doi: 10.1109/JAS.2022.105686

Michielli, N., Acharya, U. R., and Molinari, F. (2019). Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals. *Comput. Biol. Med.* 106, 71–81. doi: 10.1016/j.compbiomed.2019.01.013

Neng, W., Lu, J., and Xu, L. (2021). CCRRSleepNet: a hybrid relational inductive biases network for automatic sleep stage classification on raw single-channel EEG. *Brain Sci.* 11, 456. doi: 10.3390/brainsci11040456

Pan, Q., Brulin, D., and Campo, E. (2020). Current status and future challenges of sleep monitoring systems: systematic review. *JMIR Biomed. Eng.* 5, e20921. doi: 10.2196/20921

Perslev, M., Jensen, M., Darkner, S., Jennum, P. J., and Igel, C. (2019). "U-Time: a fully convolutional network for time series segmentation applied to sleep staging," in *Advances in Neural Information Processing Systems*, Vol. 32, eds H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Red Hook, NY: Curran Associates, Inc.), 4415?-4426.

Phan, H., Andreotti, F., Cooray, N., Chén, O. Y., and De Vos, M. (2018). Joint classification and prediction CNN framework for automatic sleep stage classification. *IEEE Trans. Biomed. Eng.* 66, 1285–1296. doi: 10.1109/TBME.2018.2872652

Rundo, J. V., and Downey, R. (2019). "Chapter 25: Polysomnography," in *Clinical Neurophysiology: Basis and Technical Aspects, volume 160 of Handbook of Clinical Neurology*, eds K. H. Levin and P. Chauvel (Amsterdam: Elsevier), 381–392.

Semyachkina-Glushkovskaya, O., Mamedova, A., Vinnik, V., Klimova, M., Saranceva, E., Ageev, V., et al. (2021). Brain mechanisms of COVID-19-sleep disorders. *Int. J. Mol. Sci.* 22:6917. doi: 10.3390/ijms22136917

Supratak, A., Dong, H., Wu, C., and Guo, Y. (2017). DeepSleepNet: a model for automatic sleep stage scoring based on raw single-channel EEG. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 1998–2008. doi: 10.1109/TNSRE.2017.2721116

Supratak, A., and Guo, Y. (2020). "TinySleepNet: an efficient deep learning model for sleep stage scoring based on raw single-channel EEG," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, 641–644. doi: 10.1109/EMBC44109.2020.9176741

Taquet, M., Geddes, J. R., Husain, M., Luciano, S., and Harrison, P. J. (2021). 6-month neurological and psychiatric outcomes in 236 379 survivors of COVID-19: a retrospective cohort study using electronic health records. *Lancet Psychiatry* 8, 416–427. doi: 10.1016/S2215-0366(21)00084-5

Van Alphen, B., Semenza, E. R., Yap, M., Van Swinderen, B., and Allada, R. (2021). A deep sleep stage in drosophila with a functional role in waste clearance. *Sci. Adv*. 7, eabc2999. doi: 10.1126/sciadv.abc2999

Wei, X., Zhang, T., Li, Y., Zhang, Y., and Wu, F. (2020). "Multi-modality cross attention network for image and sentence matching," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10938–10947. doi: 10.1109/CVPR42600.2020.01095

Yu, F., Tang, J., Yin, W., Sun, Y., Tian, H., Wu, H., et al. (2021). "ERNIE-ViL: knowledge enhanced vision-language representations through scene graphs," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 3208–3216.

Zhang, C., Yang, Z., He, X., and Deng, L. (2020a). Multimodal intelligence: representation learning, information fusion, and applications. *IEEE J. Select. Top. Signal Process*. 14, 478–493. doi: 10.1109/JSTSP.2020.2987728

Zhang, K., Su, Y., Guo, X., Qi, L., and Zhao, Z. (2020b). MU-GAN: facial attribute editing based on multi-attention mechanism. *IEEE/CAA J. Automat. Sin*. 8, 1614–1626. doi: 10.1109/JAS.2020.1003390

Zhang, Y., Xu, B., and Zhao, T. (2020c). Convolutional multi-head self-attention on memory for aspect sentiment classification. *IEEE/CAA J. Automat. Sin*. 7, 1038–1044. doi: 10.1109/JAS.2020.1003243

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 18 August 2022
DOI 10.3389/fnins.2022.974653

EDITED BY
Xiaomao Fan,
South China Normal University, China

REVIEWED BY
Ruxin Wang,
Shenzhen Institutes of Advanced Technology (CAS), China
Jingchun Liu,
Tianjin Medical University General Hospital, China

*CORRESPONDENCE
Linlin Ye
happinesslin@163.com
Weiqun Song
songwq66@126.com

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

RECEIVED 21 June 2022
ACCEPTED 29 July 2022
PUBLISHED 18 August 2022

CITATION
Cao L, Ye L, Xie H, Zhang Y and Song W (2022) Neural substrates in patients with visual-spatial neglect recovering from right-hemispheric stroke. *Front. Neurosci.* 16:974653.
doi: 10.3389/fnins.2022.974653

Neural substrates in patients with visual-spatial neglect recovering from right-hemispheric stroke

Lei Cao[1], Linlin Ye[1]*, Huanxin Xie[2], Yichen Zhang[1] and Weiqun Song[1]*

[1]Department of Rehabilitation, Xuanwu Hospital, Capital Medical University, Beijing, China,
[2]Department of Orthopedics, Beijing Rehabilitation Hospital of Capital Medical University, Beijing, China

Visual-spatial attention disorder after stroke seriously affects recovery and quality of life in stroke patients. Previous studies have shown that some patients recovery rapidly from visual-spatial neglect (VSN), but the brain networks underlying this recovery are not well understood. Using functional magnetic resonance imaging, we aimed to identify network differences between patients who rapidly recovered from VSN and those with persistent VSN. The study included 30 patients with VSN who suffered subacute stroke. Patients were examined 2–4 weeks after stroke onset and 4 weeks after the initial assessment. At the last evaluation, patients in the persistent VSN (n = 15) and rapid recovery (n = 15) groups underwent paper-and-pencil tests. We defined the bilateral frontal eye fields, bilateral intraparietal sulcus in the dorsal attention network, and right temporoparietal junction and ventral frontal cortex areas in the ventral attention network as regions of interest (ROI) and measured whole-brain ROI-based functional connectivity (FC) and amplitude of low-frequency fluctuations (ALFF) in subacute right-hemisphere stroke patients. VSN recovery was associated with changes in the activation of multiple bilateral attentional brain regions. Specifically, persistent VSN was associated with lower FC in the right superior frontal gyrus, right inferior temporal gyrus, right medial orbitofrontal cortex, left precuneus, right inferior parietal gyrus, right medial frontal gyrus, right rectus gyrus, left superior frontal gyrus, left middle cingulate gyrus, right superior temporal pole, right postcentral gyrus, and right posterior cingulate gyrus compared to that in those with rapid recovery, whereas ALFF in the left cerebellum were decreased in patients with persistent VSN. Our results demonstrate that the DAN rather than the VAN, plays a more important role in recovery from VSN, and that the cerebellum is involved in recovery. We believe that our results supplement those of previous studies on recovery from VSN.

KEYWORDS

visual-spatial neglect, recovery, dorsal attention network, functional connectivity, ALFF

Introduction

Visual-spatial neglect (VSN) is one of the most common cognitive impairments after stroke, particularly in the right hemisphere. VSN occurs in more than half of right hemisphere stroke survivors (Esposito et al., 2020). Unfortunately, in approximately 40% of patients, VSN persists to 1 year after onset (Nijboer et al., 2013) and is associated with poor functional outcome (Chen et al., 2015).

The mechanisms underlying this recovery remain largely unknown. Functional magnetic resonance imaging (fMRI) studies have revealed global activation changes in functional organization far beyond the lesions in VSN patients after stroke (Corbetta et al., 2005). In fact, attentional networks, dorsal attention networks (DAN), and ventral attention networks (VAN) have been identified as crucial components of VSN recovery (Corbetta and Shulman, 2011). A few VSN studies have shown that interhemispheric and intrahemispheric functional connections between attention networks affect recovery from VSN. Baldassarre et al. (2014) highlighted that large-scale changes in network interactions following focal injury are associated with decreased interhemispheric functional connectivity (FC) and the patterns of abnormal functional connectivity between VAN and DAN. Further studies have shown that attention deficit following stroke is significantly more correlated with interhemispheric FC in the DAN and proposed interventions that restore normal patterns of resting FC may be associated with good recovery (Baldassarre et al., 2016). Two important longitudinal studies have demonstrated that improvement of VSN was correlated with an increase in previously depressed interhemispheric FC across attention and network normalization of interhemispheric regions in multiple networks, predominantly the left functional homologs (Ramsey et al., 2016; Umarova et al., 2016).

Recent research extracted the amplitude of low-frequency fluctuations (ALFF) from each region of interest (ROI) as a marker of regional spontaneous neuronal activity in acute right-hemisphere stroke patients with left hemispatial neglect, which can complement the fact that FC analysis cannot offer local dysfunction in a specific area of the network. They also confirmed a reduction in spontaneous neuronal activity in the superior parietal lobule as a consequence of VSN. Intervention at the superior parietal lobule may improve left VSN behavior in stroke patients (Machner et al., 2020).

Previous studies have shown that some patients recover rapidly from VSN (Ye et al., 2020). However, these studies did not consider the effects of the DAN and VAN. We investigated attention network differences between good and poor recovery from VSN to identify specific features of favorable recovery from VSN after a right hemisphere stroke by using two parameters of resting-state fMRI (rs-fMRI): ALFF and FC.

Patients and methods

Patient group and clinical details

We retrospectively analyzed data from 30 patients who were admitted to the rehabilitation department of Xuanwu Hospital with a subacute stroke and tested within 2-4 weeks of stroke onset between July 2017 and October 2018. The patients were diagnosed with first-onset right-hemisphere stroke and VSN. The inclusion criteria were: (1) right-handedness, (2) age $\geq$ 18 years, and (3) availability of informed consent. The exclusion criteria were: (1) non-cooperation with paper-and-pencil tests, (2) previous history of a psychiatric or neurological disorder, (3) claustrophobia, (4) metal implants such as artificial cochlea and cardiac pacemaker; and (5) bilateral-hemisphere or previous brain lesions. All patients received upper- and lower-extremity exercise therapy for 1 h twice daily for 2 weeks. No specific intervention was provided for the VSN. All patients completed paper-and-pencil tests and underwent MRI before exercise therapy.

This study was approved by the local ethics committee and written informed consent was obtained from all participants.

Neuropsychological assessment

Standardized paper-and-pencil tests were used to evaluate the severity of the VSN. These tests included the line-bisection, line-cancelation, star-cancelation, clock-drawing, and sentence-reading tests. All tests were conducted on a horizontally placed 295 × 210 mm A4 sheet of paper. Behavioral scores were calculated using previously described methods. In the line-bisection task, patients were instructed to mark the midpoints of five horizontal lines of different lengths (between 80 and 160 mm) distributed on the paper. Rightward or leftward deviations from the real midpoint were measured, and a deviation $>12\%$ was considered pathological neglect. In the line-cancelation task, patients were instructed to mark 30 line symbols, and a difference between left- and right-sided omissions of three or more was considered pathological neglect. In the star-cancelation task, patients were instructed to mark all small stars that were symmetrically distributed among the distractors on the test paper (27 in the left field, two in the middle field, and 27 in the right field). A difference between left- and right-sided omissions of more than five was considered pathological neglect. In the clock-drawing task, a clock with a random time was presented on paper, and participants were instructed to copy a duplicate clock on the test paper; omission of the left side of the clock compared with the right side was considered pathological neglect. In the sentence-reading task, participants were instructed to read an article aloud in three columns (left, center, and right). Omission of words in one or more sentences was considered pathological neglect. We calculated the scores by considering the total number of targets omitted and left-right differences.

Subjects were diagnosed with VSN when the results of at least two of the paper-and-pencil tests were positive. Behavioral results were independently evaluated by two neurologists.

Paper-and-pencil tests were performed in the early subacute phase (2–4 weeks after stroke) and 4 weeks after the first assessment. MRI was performed in the same day after first assessment. At four weeks after the second assessment, 15 patients still had VSN and were assigned to the persistent VSN group, while the remaining 15 patients were free of VSN, as judged by all behavioral tests, and were assigned to the rapid recovery group. We carefully checked the clinical profiles of all the patients without neglect and found no signs of neglect. Subsequently, a retrospective analysis was performed.

MRI acquisition

Data were obtained using a GE Signa 3.0T scanner. Foam pads were used to prevent head movement. Functional images were obtained using an echo-planar imaging sequence with the following parameters: 33 axial slices, thickness/gap = 3.5/0 mm, in-plane resolution = 64 × 64, repetition time = 2000 ms, echo time = 30 ms, flip angle = 90°, and field of view = 212 × 212 mm^2. The resting-state MRI sessions lasted for 8 minutes. Patients were instructed to hold still, plug their ears with sponge earplugs, not think systematically, and not fall asleep. In addition, a T1-weighted sagittal three-dimensional magnetization-prepared rapid gradient echo sequence was acquired using the following parameters:144 slices, repetition time = 2,300 ms, echo time = 3.39 ms, slice thickness = 1 mm, flip angle = 7°, inversion time = 1100 ms, field of view = 200 × 256 mm^2, and in-plane resolution = 200 × 256. T1-weighted MRI protocols had the following parameters: 32 slices, repetition time = 3823.5 ms, echo time = 24 ms, flip angle = 111°, matrix 256 × 288, 32 slices. The standard slice thickness was 3.0 mm.

Regions of interest selection

Six ROI were selected based on previous research (Farrant and Uddin, 2015). In the DAN, the ROI were located in the right frontal eye field (rFEF) (Montreal Neurological Institute [MNI] coordinates: $x = 28$, $y = -8$, $z = 52$), left FEF (lFEF) (MNI coordinates: $x = -28$, $y = -8$, $z = 52$), right intraparietal sulcus (rIPS) (MNI coordinates: $x = 21$, $y = -58$, $z = 53$), left IPS (lIPS) (MNI coordinates: $x = -28$, $y = -56$, $z = 44$), right temporoparietal junction (TPJ) (MNI coordinates: $x = 60$, $y = -48$, $z = 22$), and right ventral frontal cortex (VFC) (MNI coordinates: $x = 42$, $y = 20$, $z = -6$) for ventral attention networks.

Imaging data analysis

The fMRI data were processed and analyzed performed using the Data Processing Assistant of the rs-fMRI software (DPARSF Advanced Edition V5.2)[1]. The first 10 volumes for each participant were discarded. All images were time-shifted, such that the slices were temporally aligned. The images were then realigned, and all participants moved no more than 3 mm in translational or 3° in rotational dimensions. The images were then co-registered with anatomical images, which were segmented into gray and white matter. Anatomical images were obtained using Diffeomorphic Anatomical Registration Exponentiated Lie Algebra (DARTEL). First, a sample-specific template was generated from the T1-weighted images. Second, the individual anatomical images were normalized non-linearly to the template, followed by linear registration to the MNI template. The images were smoothed using a Gaussian filter with a full width at half maximum of 4 mm. Finally, nuisance signals, including 24 head-motion parameters, cerebrospinal fluid signals, and white matter signals, were regressed from the MRI data.

Further preprocessing, including removal of linear trends, temporal band-pass filtering (0.01–0.1 Hz), and ALFF, was performed using the Data Processing Assistant of rs-fMRI software (DPARSF Advanced Edition V5.2; See Text Footnote 1). All ALFF map results were converted into z-maps.

To perform the FC analyses, time series from the resting-state scan were extracted for the subject-specific ROI in the rFEF, lFEF, rIPS, and lIPS for DAN, and right TPJ and right VFC for VAN by averaging the time series of all voxels in the spherical ROI (radius = 6 mm). Whole-brain voxel-based FC analysis was used to calculate the FC strength between each voxel in each subject and the ROI and convert it into z-maps.

Statistical analyses

SPSS (version 22.0; IBM Corp., Armonk, NY, United States) was used for all statistical analyses. Demographic characteristics, including symptom severity, were analyzed using descriptive statistics (mean differences ± standard deviation ($x \pm s$) for each group. Sex and type of stroke were compared using the $\chi 2$ test, and other quantitative variables were compared using the independent-sample *t* test. Statistical significance was set at $p \leq 0.05$.

Group-level statistical analyses were performed using the statistical analysis model in DPARSF. A two-sample t-test was conducted on the individual z-ALFF and z-ROI FC maps of the two groups with a small volume correction for the one-sample

---

[1] http://www.rfmri.org/DPARSF

result masks. Multiple-comparison corrections were performed using the Gaussian random field correction. Significant between-group differences met the criteria of corrected $p < 0.01$ for voxel level and $p < 0.05$ for cluster size level.

Results

Patient classification and demographics

The average age of the persistent VSN patients and the rapid recovery patients were 57.27 ± 10.93 and 56.20 ± 10.93 years, respectively (the persistent VSN group: three women, 12 men, age 32–73 years; the rapid recovery group: four women, 11 men, age 37–71 years). There were no significant differences in age, sex, type of stroke, years of education, initial paper-and-pencil test scores, or clinical course following stroke between the two groups ($p > 0.05$). The demographic and clinical characteristics of the patients are presented in **Table 1**. The overlapped lesion plots of the patients are presented in **Figure 1**.

Neuropsychological performance

The behavioral scores in the early subacute phase showed no significant difference between the rapid recovery and persistent VSN patients (Line bisection $t = -0.005$, $p = 0.996$; Line cancelation $t = 0.688$, $p = 0.498$; Star cancelation $t = 0.044$, $p = 0.965$; Sentence reading $t = 0.417$, $p = 0.680$). The behavioral scores of the two groups are presented in **Figure 2**.

Differences in whole-brain functional connectivity of dorsal attention networks nodes

In the persistent VSN group, the rFEF showed strong FC with the right precentral gyrus, left precuneus, left precentral gyrus, left middle cingulate gyrus, left supplementary motor area, and left cerebellum. In the rapid recovery group, the rFEF showed strong FC with the right supplementary motor area, right superior frontal gyrus, right precentral gyrus, left lingual gyrus, left cerebellum, left medial orbitofrontal cortex, right middle temporal gyrus, and left middle frontal gyrus (**Figure 3A**).

In the persistent VSN group, the lFEF showed strong FC with the left superior frontal gyrus, left triangular part of the inferior frontal gyrus, and left putamen. In the rapid recovery group, the lFEF showed strong FC with the left precentral gyrus, right cerebellum, left inferior temporal gyrus, and left cerebellum (**Figure 3B**).

In the persistent VSN group, the rIPS showed strong FC with the right superior parietal gyrus, left lingual gyrus, left superior marginal gyrus, right supplementary motor area, left middle cingulate gyrus, right middle temporal pole, right superior marginal gyrus, and right fusiform gyrus. In the rapid recovery group, the rIPS showed strong FC with the right superior parietal gyrus, left thalamus, right superior frontal gyrus, right middle frontal gyrus, left medial superior frontal gyrus, left cerebellum, right cerebellum, right middle temporal gyrus, right superior frontal gyrus, left triangular part of the inferior frontal gyrus, left middle temporal gyrus, and left opercular part of the inferior frontal gyrus (**Figure 4A**).

In the persistent VSN group, the lIPS showed strong FC with the left superior parietal gyrus and left cerebellum. In the rapid recovery group, the lIPS showed strong FC with the left inferior parietal gyrus and right rolandic operculum (**Figure 4B**).

The FC between the DAN and voxel-based whole-brain of the rapid recovery group was significantly increased compared to that of the persistent VSN group ($p < 0.01$). The FC between the rFEF, lFEF, rIPS, lIPS, and the whole brain is shown in **Figure 3**. The rFEF ROI showed significantly stronger FC with the right superior frontal gyrus, right inferior temporal gyrus, right medial orbitofrontal cortex, left precuneus, and right inferior parietal gyrus in the rapid recovery group than in the persistent VSN group. The lFEF ROI showed significantly stronger FC with the right superior frontal gyrus, right medial frontal gyrus, right rectus gyrus, and left superior frontal gyrus in the rapid recovery group than in the persistent VSN group. The rIPS ROI showed significantly stronger FC with the left middle cingulate gyrus in the rapid recovery group than in the persistent VSN group. The lIPS ROI showed significantly stronger FC with the right superior temporal pole, right postcentral gyrus, and right posterior cingulate gyrus in the rapid recovery group than in the persistent VSN group (**Table 2** and **Figures 3**, **4**).

Differences in whole-brain functional connectivity of ventral attention networks nodes

In both the persistent VSN and rapid recovery groups, the right TPJ showed strong FC in the right superior temporal gyrus.

In both the persistent VSN and rapid recovery groups, the right VFC showed strong FC in the right insula.

In terms of FC with other VAN regions, no group differences were found in VAN (TPJ/VFC) connections. In other words, the persistent VSN and rapid recovery groups did not differ in FC strength between VAN and whole-brain voxel-based functional connections.

TABLE 1 Demographic data and clinical data.

Patient ID	Age	Sex	Time science stroke (days)	Type of stroke	Years of education	Line bisection (deviation%)	Line cancelation (all omissions)	Star cancelation (all omissions)	Clock drawing	Sentence reading omissions
P1	32	M	28	CH	18	54.8	28	38	+	5
P2	69	M	28	CI	15	39.7	15	38	+	4
P3	54	M	25	CI	12	35.66	26	48	+	9
P4	51	M	20	CI	14	64.95	8	36	+	6
P5	57	M	21	CI	11	17.26	5	14	+	2
P6	51	M	27	CI	12	14.72	6	24	+	2
P7	57	F	20	CH	16	78.04	24	48	+	13
P8	70	M	20	CI	9	52.95	9	13	+	2
P9	62	M	14	CI	9	69.97	17	48	+	7
P10	55	M	19	CI	15	20.95	6	42	+	4
P11	41	F	21	CI	18	21.46	6	26	+	4
P12	73	M	23	CI	8	82.69	15	27	+	9
P13	65	F	27	CI	12	42.37	4	22	+	2
P14	64	M	24	CI	9	50	23	46	+	7
P15	58	M	14	CI	12	32.75	5	9	+	2
R1	37	F	26	CI	18	69.46	21	48	+	4
R2	53	M	25	CH	12	42.34	14	46	+	7
R3	71	M	26	CI	6	56.07	18	34	+	4
R4	50	M	21	CI	3	69.96	6	25	+	6
R5	65	M	19	CI	9	33	15	40	+	5
R6	47	M	22	CI	12	25.4	14	25	+	2
R7	42	M	22	CI	16	68.37	12	30	+	6
R8	52	M	17	CI	15	38.26	4	25	+	4
R9	52	M	27	CI	14	62.96	10	45	+	13
R10	47	F	18	CI	19	28.7	19	34	+	2
R11	60	M	26	CI	9	75.23	10	17	+	6
R12	67	M	26	CH	12	42.27	7	33	+	4
R13	72	F	20	CI	6	20.2	6	16	+	2
R14	59	F	21	CI	12	16.67	9	44	+	4
R15	69	M	20	CI	9	30	5	14	+	2

P1-P15: patients with persistent VSN; R1-R15: patients with rapid recovery VSN; M: male; F: female; CH: cerebral hemorrhage; CI: cerebral ischemia.

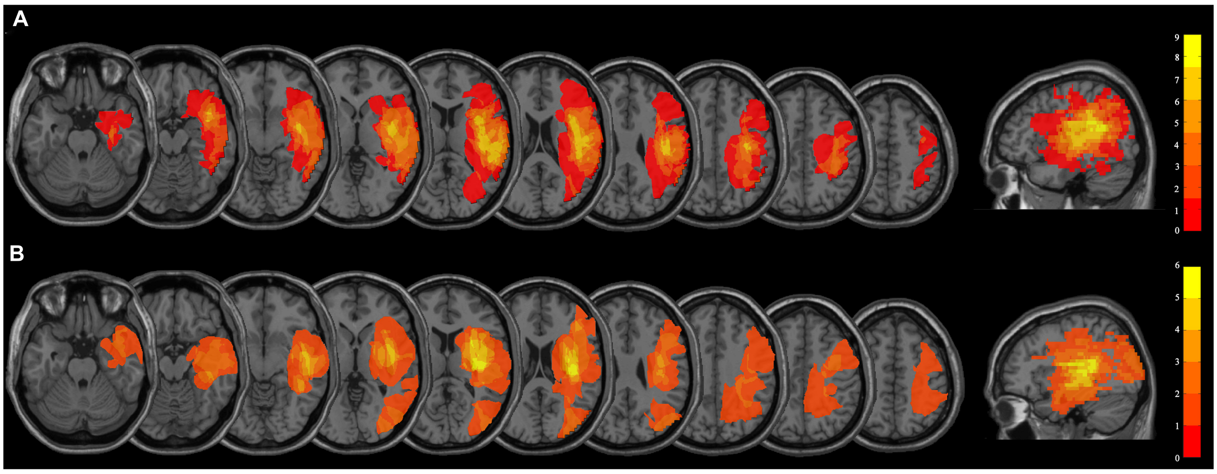


FIGURE 1
(A) Overlapped lesion plots of the patients with persistent VSN. **(B)** Overlapped lesion plots of the patients with rapid recovery VSN.

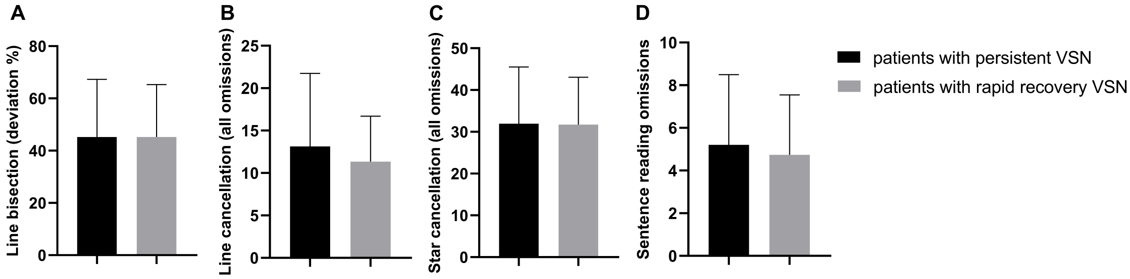


FIGURE 2
The behavioral scores in two groups, all $p > 0.05$. **(A)** Line bisection (deviation %); **(B)** Line cancellation (all omissions); **(C)** Star cancellation (all omissions); **(D)** Sentence reading omissions.

Amplitude of low-frequency fluctuations

At the group level (**Figure 5**), the z-ALFF value did not differ significantly between the persistent VSN and rapid recovery groups for most ROI. Only the mean z-ALFF value in the left cerebellum anterior lobe was significantly reduced in the persistent VSN group compared with that in the rapid recovery group.

Discussion

VSN is common following stroke, but the neural substrates for recovery are unclear. Our results showed that the interhemispheric and intrahemispheric FC of the DAN to the whole brain was reduced in patients with persistent VSN compared to that in patients with rapid recovery. The results provide novel insights into the mechanisms underlying recovery; that is, the FC of the DAN, but not the VAN, key nodes play a crucial role in recovery from VSN. In addition, local reduction in spontaneous neuronal activity in the cerebellum is associated with poor recovery.

Recovery is associated with changes in FC in the DAN. Consistent with previous research, changes in interhemispheric FC are sensitive markers related to recovery, especially in the DAN (Ramsey et al., 2016; Umarova et al., 2016). In this study, we found that the ipsilesional DAN was more related to recovery by an increase in the interhemispheric FC of the key nodes in the ipsilesional frontal, parietal, and temporal regions, whereas the contralesional DAN was more related to the key nodes of the lesionlateral hemisphere. This indicates that both the ipsilesional and contralesional hemispheres are involved in compensation of the attention network after injury. This may involve reorganization and integration of the attention network. Umarova et al. conducted a longitudinal study on patients with VSN and found that transneuronal changes in the contralesional frontoparietal and bilateral occipital connections were distinctly related to an unrecovered VSN in patients with chronic stroke.

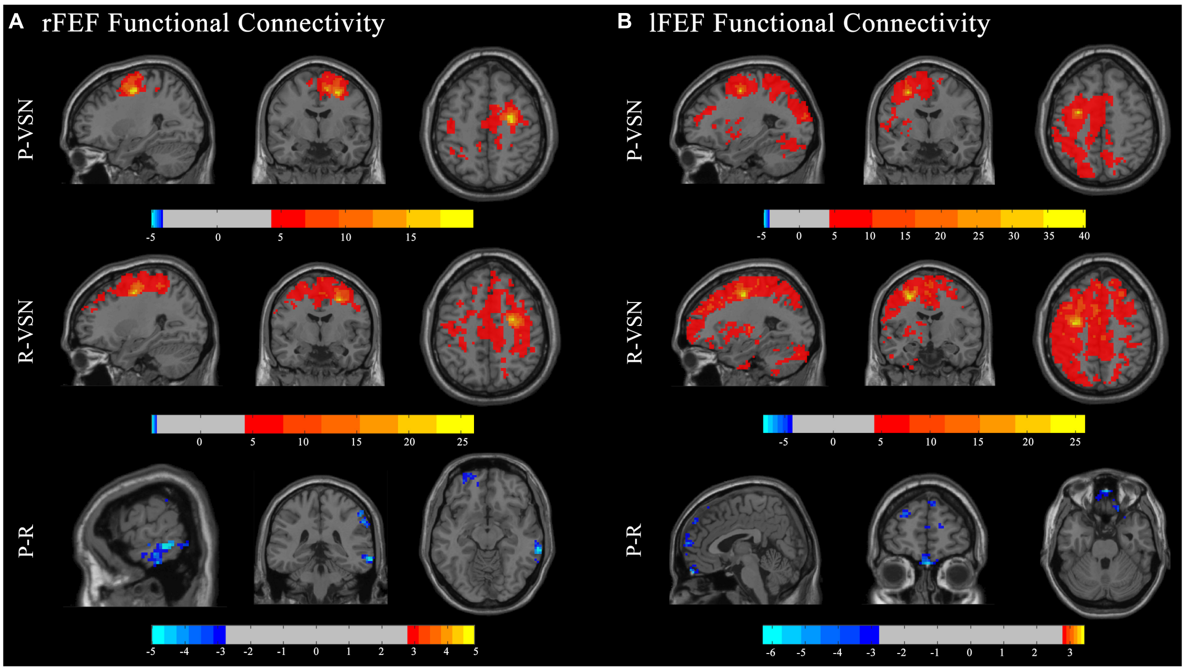


FIGURE 3

Functional connectivity (FC) maps. Individual and group comparisons of resting state functional connectivity of seed ROIs rFEF **(A)** and lFEF **(B)**. First panel show individual one sample *t*-test for persistent VSN(P-VSN) ($p < 0.001$). Second panel show individual one sample *t*-test for rapid recovery VSN(R-VSN) ($p < 0.001$). Remaining panels show two-sample t-tests comparing persistent VSN with rapid recovery VSN(P-R) ($p < 0.01$).

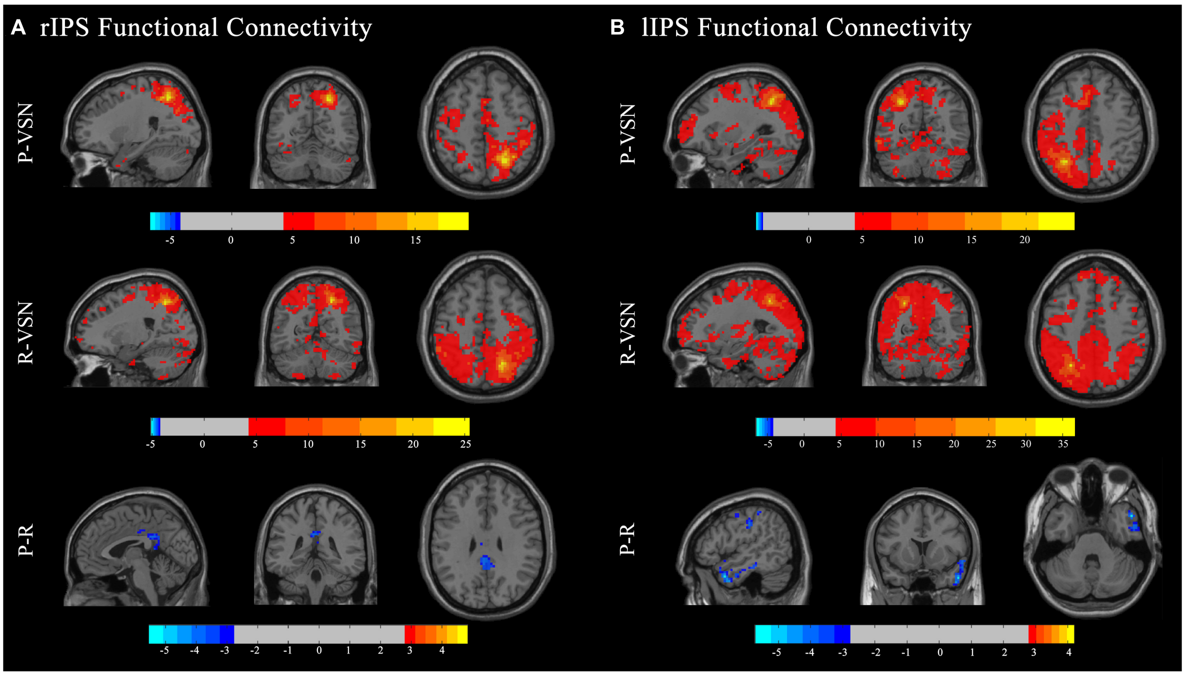


FIGURE 4

Functional connectivity (FC) maps. Individual and group comparisons of resting state functional connectivity of seed ROIs rIPS **(A)** and lIPS **(B)**. First panel show individual one sample t-test for persistent VSN(P-VSN) ($p < 0.001$). Second panel show individual one sample t-test for rapid recovery VSN(R-VSN) ($p < 0.001$). Remaining panels show two-sample t-tests comparing persistent VSN with rapid recovery VSN(P-R) ($p < 0.01$).

TABLE 2 Group differences in functional connectivity of DAN and VAN ROIs.

Seed	Functional connectivity	MNI coordinates			*Z*-score	Cluster size
		X	*Y*	*Z*		(voxels)
rFEF	right superior frontal gyrus	15	48	33	−4.92	456
	right inferior temporal gyrus	66	−36	−12	−5.01	211
	right medial orbitofrontal cortex	−15	69	−3	−4.59	186
	left precuneus	−6	−66	39	−3.55	152
	right inferior parietal gyrus	60	−33	39	−4.87	114
lFEF	right superior frontal gyrus	6	54	39	−4.52	206
	right medial frontal gyrus	−21	60	27	−5.36	204
	right rectus gyrus	3	57	−27	−5.53	116
	left superior frontal gyrus	−9	42	54	−4.45	110
rIPS	left middle cingulate gyrus	−3	39	30	−3.94	104
lIPS	right superior temporal pole	51	12	−33	−5.72	428
	right postcentral gyrus	54	21	39	−4.53	376
	right posterior cingulate gyrus	0	−54	24	−4.13	117

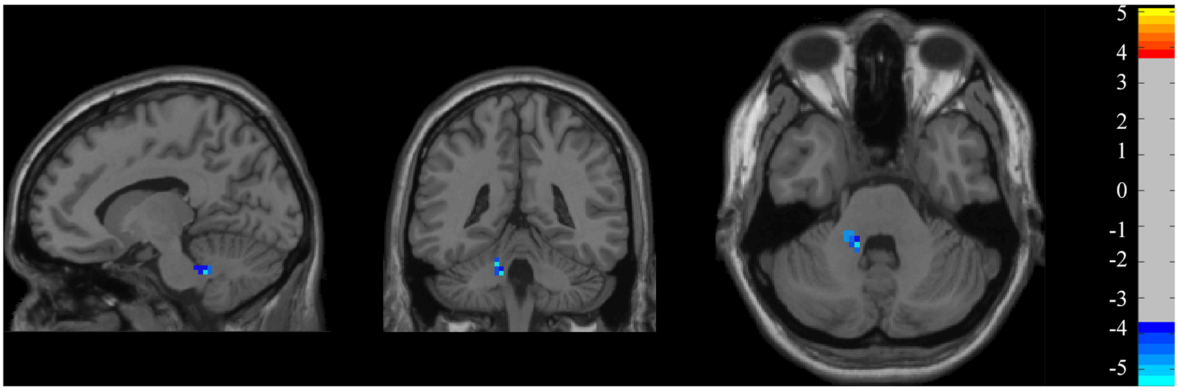


FIGURE 5
ALFF maps. ALFF maps show two-sample *t*-tests comparing persistent VSN with rapid recovery VSN(P-R) ($p < 0.001$).

This revealed a large-scale structural reorganization of the visual-spatial attention network after stroke, especially in the trans-hemisphere (Umarova et al., 2017). Lunven et al. believed that persistent neglect is due to disconnection between the DAN and VAN. This emphasizes the importance of network-based integration (Lunven et al., 2015).

Large-scale reorganization of functional connections in the whole-brain network structure has been previously reported. Animal studies have shown that when one hemisphere is damaged, activation of the entire attention network decreases, in addition to the decrease in activation of the ipsilesional attention network. Synaptic loss occurred 2–4 days after stroke, the normal network activation mode was disrupted, and the entire attention network was in an overexcited state 7 days after stroke, and characterized by the expansion of excitatory neurites and an increase in synapses. This phenomenon exists not only around the lesion but also in the contralesional hemisphere (Deuel and Collins, 1983; Gonzalez et al., 2004). A recent study found that axonal and retrograde degeneration occurs not only occur around the lesion but also in distant regions that are not directly connected to the lesion by transneuronal remodeling (Fornito et al., 2015). In addition, the lesions in our patients were located within the territory of the right middle cerebral artery. According to some studies, hemispheric differences occur in different hemispheres after an acute middle cerebral artery infarction. Compensatory activation occurs in related brain areas of the dominant hemisphere after acute infarction in the middle cerebral artery of the non-dominant hemisphere (Gao et al., 2021).

In the current study, we also found that patients with persistent VSN showed weaker FC of the DAN to the whole brain, particularly in the frontal region, as revealed by group differences in the whole-brain FC of key DAN nodes. Previous studies have indicated an important role of the frontal lobe in the

DAN. The contralesional prefrontal cortex supports to residual ipsilesional attention centers to accelerate restoration or the reorganization of the perilesional cortex (Umarova et al., 2016). According to the degree of injury to the patient, the prefrontal cortex produces different advanced controls over movement and sensation systems (Liu et al., 2016). In stroke patients with mild dysfunction, rs-fMRI has shown that a stronger FC between the prefrontal lobe and motor cortex is associated with better recovery (Lam et al., 2018). However, in patients with severe dysfunction after stroke, the prefrontal cortex may play a greater role in functional recovery, and FC between the prefrontal sensorimotor cortex was significantly correlated with recovery (Yin et al., 2012). In unilateral hemispheric prefrontal lobe injury in macaque monkeys, FC between the contralesional prefrontal cortex and ipsilesional parietal cortex was related to recovery (Adam et al., 2020). Activation of the contralesional prefrontal cortex is a compensatory strategy underlying cortical mechanisms during the recovery process after VSN (Takamura et al., 2016).

In addition to conventional FC analyses, we extracted the ALFF from each ROI as a marker of spontaneous regional neuronal activity. The ALFF is a low-frequency fluctuation index that improves sensitivity and specificity in the detection of spontaneous brain activitiy (Zou et al., 2008). Interestingly, we found that the ALFF values in the left cerebellum anterior lobe were lower in the persistent VSN group than in the rapid recovery group. This indicates that the cerebellum has a positive effect on recovery and may compensate for the VSN by affecting the attention network. Inconsistent with previous studies by Machner et al., we found that the ALFF of fMRI in the cerebellum of patients with persistent VSN was lower than that of patients with rapid recovery. However, the reduced ALFF also reflects local disruption of long-range neuronal network communication in VSN patients (Machner et al., 2020).

These findings do not contradict the previous proposal of a crucial role for the DAN in VSN. However, the functional role of cerebellar structures in cognitive networks remains poorly understood. In fact, an article published in *Science* in 1997 was the first fMRI study propose the role of the cerebellum in attention processing (Allen et al., 1997). However, in the following 20 years, the role of the cerebellum in motor function gradually came into view, and its role in cognition was rarely mentioned. Buckner et al. identified a region spanning cerebellar lobules VIIb and VIIIa that exhibited connectivity with the cortical DAN (Buckner et al., 2011). In electrophysiological tests of patients with cerebellar injury, the lateral CrusI and CrusII of the cerebellum were associated with the attention network of the frontal parietal lobe (Striemer et al., 2015). A few months later, Brissenden et al. found that a region located in the posterior cerebellar lobe (lobules VIIb and VIIIa) showed strong activation under the attention multi-objective paradigm through visual attention-related tasks, which provided evidence that the DAN extended functionally to the cerebellum (Brissenden et al., 2016). Subsequently, this team demonstrated that neurons in the VIIb/VIIIa area of the cerebellum generates stronger recruitment during spatial coding tasks and were strongly activated under visual working memory load. This performance was mirrored by the corresponding areas of the cerebral hemisphere, which further proved that there was a highly specific cortical cerebellar network involved in attention function (Brissenden et al., 2018). Furthermore, a recent study showed that the cerebellum, as an attention structure across the cortex and subcortex, was strongly activated during the attention paradigm, which provides new evidence for the cerebellum as an important node of the DAN (Brissenden et al., 2016). There are more neurons in the intact cerebellum than in the cerebral cortex that may participate in the DAN and compensate for recovery through the cerebellum-to-cortex circuit. Consistent with our results, all the patients with good outcomes showed cerebellar activation (Lasek-Bal et al., 2018). The results of these studies also support the idea that neurological deficits are compensated by engagement of larger areas of the cerebral cortex.

The potential limitations of this study should be considered. First, this was a retrospective study that compared differences in FC between patients with different outcomes. However, we did not include right brain lesions without VSN. Second, due to the high requirements of rs-fMRI for patients, we included only patients who could cooperate with the examination. Therefore, we were unable to assess the attention network of patients with a more serious disability who could not cooperate in the examination. Third, our study focused on patients with VSN in the early subacute stage after stroke. There has been no follow-up of these patients, and it is unknown whether patients with persistent VSN also experience neglect in the chronic phase. Finally, patients with cerebral hemorrhage and cerebral infarction were included in this study. To minimize the impact of cerebral hemorrhage on the results, we included examinations of patients 2–4 weeks after stroke. Cerebral hemorrhage was resolved in four patients, and there was no obvious perilesional edema. We also believe that it would be better to unify the types of stroke.

Conclusion

Persistent VSN patients with stroke show interhemispheric and intrahemispheric FC of the DAN to the whole brain, and the ALFF value in the left cerebellum anterior lobe was reduced compared to that in those with rapid recovery. Our results demonstrate that the DAN rather than the VAN, plays a more important role in recovery from VSN, and that the cerebellum is involved in recovery. Our findings provide activation in

DAN between key notes as important evidence for recovery and facilitate the design of therapeutic approaches such as non-invasive brain stimulation to ensure better recovery of VSN and development of newer interventions after stroke.

Data availability statement

The original contributions presented in this study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics committee of Xuanwu Hospital. The patients/participants provided their written informed consent to participate in this study.

Author contributions

LC: drafting and revising the manuscript, study concept and design, and analysis and interpretation of data. LY: study concept and design, acquisition of data, revising the manuscript, and obtaining funds. HX: analysis and interpretation of data. YZ: analysis of data and acquisition of MRI data. WS: study concept and design, analysis and interpretation of data, and study supervision. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 82002386) and the Beijing Municipal Administration of Hospitals Incubating Program (code: PX2021037). The funding sources had no involvement in the study design, in the collection, analysis and interpretation of data, in the writing of the report, or in the decision to submit this article for publication.

Conflict of interest

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

Publisher's note

References

Adam, R., Johnston, K., Menon, R. S., and Everling, S. (2020). Functional reorganization during the recovery of contralesional target selection deficits after prefrontal cortex lesions in macaque monkeys. *Neuroimage* 207:116339. doi: 10.1016/j.neuroimage.2019.116339

Allen, G., Buxton, R. B., Wong, E. C., and Courchesne, E. (1997). Attentional activation of the cerebellum independent of motor involvement. *Science* 275, 1940–1943. doi: 10.1126/science.275.5308.1940

Baldassarre, A., Ramsey, L., Hacker, C. L., Callejas, A., Astafiev, S. V., Metcalf, N. V., et al. (2014). Large-scale changes in network interactions as a physiological signature of spatial neglect. *Brain* 137(Pt 12), 3267–3283. doi: 10.1093/brain/awu297

Baldassarre, A., Ramsey, L., Rengachary, J., Zinn, K., Siegel, J. S., Metcalf, N. V., et al. (2016). Dissociated functional connectivity profiles for motor and attention deficits in acute right-hemisphere stroke. *Brain* 139(Pt 7), 2024–2038. doi: 10.1093/brain/aww107

Brissenden, J. A., Levin, E. J., Osher, D. E., Halko, M. A., and Somers, D. C. (2016). Functional Evidence for a cerebellar node of the dorsal attention network. *J. Neurosci.* 36, 6083–6096. doi: 10.1523/JNEUROSCI.0344-16.2016

Brissenden, J. A., Tobyne, S. M., Osher, D. E., Levin, E. J., Halko, M. A., and Somers, D. C. (2018). Topographic cortico-cerebellar networks revealed by visual attention and working memory. *Curr. Biol.* 336:e3365. doi: 10.1016/j.cub.2018.08.059

Buckner, R. L., Krienen, F. M., Castellanos, A., Diaz, J. C., and Yeo, B. T. (2011). The organization of the human cerebellum estimated by intrinsic functional connectivity. *J. Neurophysiol.* 106, 2322–2345. doi: 10.1152/jn.00339.2011

Chen, P., Hreha, K., Kong, Y., and Barrett, A. M. (2015). Impact of spatial neglect on stroke rehabilitation: Evidence from the setting of an inpatient rehabilitation facility. *Arch. Phys. Med. Rehabil.* 96, 1458–1466. doi: 10.1016/j.apmr.2015.03.019

Corbetta, M., and Shulman, G. L. (2011). Spatial neglect and attention networks. *Annu. Rev. Neurosci.* 34, 569–599. doi: 10.1146/annurev-neuro-061010-113731

Corbetta, M., Kincade, M. J., Lewis, C., Snyder, A. Z., and Sapir, A. (2005). Neural basis and recovery of spatial attention deficits in spatial neglect. *Nat. Neurosci.* 8, 1603–1610. doi: 10.1038/nn1574

Deuel, R. K., and Collins, R. C. (1983). Recovery from unilateral neglect. *Exp. Neurol.* 81, 733–748. doi: 10.1016/0014-4886(83)90340-0

Esposito, E., Shekhtman, G., and Chen, P. (2020). Prevalence of spatial neglect post-stroke: A systematic review. *Ann. Phys. Rehabil. Med.* 64:101459. doi: 10.1016/j.rehab.2020.10.010

Farrant, K., and Uddin, L. Q. (2015). Asymmetric development of dorsal and ventral attention networks in the human brain. *Dev. Cogn. Neurosci.* 12, 165–174. doi: 10.1016/j.dcn.2015.02.001

Fornito, A., Zalesky, A., and Breakspear, M. (2015). The connectomics of brain disorders. *Nat. Rev. Neurosci.* 16, 159–172. doi: 10.1038/nrn3901

Gao, J., Yang, C., Li, Q., Chen, L., Jiang, Y., Liu, S., et al. (2021). Hemispheric difference of regional brain function exists in patients with acute stroke in different cerebral hemispheres: A resting-state fMRI study. *Front. Aging Neurosci.* 13:691518. doi: 10.3389/fnagi.2021.691518

Gonzalez, C. L., Gharbawie, O. A., Williams, P. T., Kleim, J. A., Kolb, B., and Whishaw, I. Q. (2004). Evidence for bilateral control of skilled movements: Ipsilateral skilled forelimb reaching deficits and functional recovery in rats follow motor cortex and lateral frontal cortex lesions. *Eur. J. Neurosci.* 20, 3442–3452. doi: 10.1111/j.1460-9568.2004.03751.x

Lam, T. K., Dawson, D. R., Honjo, K., Ross, B., Binns, M. A., Stuss, D. T., et al. (2018). Neural coupling between contralesional motor and frontoparietal networks correlates with motor ability in individuals with chronic stroke. *J. Neurol. Sci.* 384, 21–29. doi: 10.1016/j.jns.2017.11.007

Lasek-Bal, A., Kidon, J., Blaszczyszyn, M., Stasiow, B., and Zak, A. (2018). BOLD fMRI signal in stroke patients and its importance for prognosis in the subacute disease period – Preliminary report. *Neurol. Neurochir. Pol.* 52, 341–346. doi: 10.1016/j.pjnns.2017.12.006

Liu, H., Tian, T., Qin, W., Li, K., and Yu, C. (2016). Contrasting evolutionary patterns of functional connectivity in sensorimotor and cognitive regions after stroke. *Front. Behav. Neurosci.* 10:72. doi: 10.3389/fnbeh.2016.00072

Lunven, M., Thiebaut De Schotten, M., Bourlon, C., Duret, C., Migliaccio, R., Rode, G., et al. (2015). White matter lesional predictors of chronic visual neglect: A longitudinal study. *Brain* 138(Pt 3), 746–760. doi: 10.1093/brain/awu389

Machner, B., von der Gablentz, J., Gottlich, M., Heide, W., Helmchen, C., Sprenger, A., et al. (2020). Behavioral deficits in left hemispatial neglect are related to a reduction of spontaneous neuronal activity in the right superior parietal lobule. *Neuropsychologia* 138:107356. doi: 10.1016/j.neuropsychologia.2020.107356

Nijboer, T. C., Kollen, B. J., and Kwakkel, G. (2013). Time course of visuospatial neglect early after stroke: A longitudinal cohort study. *Cortex* 49, 2021–2027. doi: 10.1016/j.cortex.2012.11.006

Ramsey, L. E., Siegel, J. S., Baldassarre, A., Metcalf, N. V., Zinn, K., Shulman, G. L., et al. (2016). Normalization of network connectivity in hemispatial neglect recovery. *Ann. Neurol.* 80, 127–141. doi: 10.1002/ana.24690

Striemer, C. L., Cantelmi, D., Cusimano, M. D., Danckert, J. A., and Schweizer, T. A. (2015). Deficits in reflexive covert attention following cerebellar injury. *Front. Hum. Neurosci.* 9:428. doi: 10.3389/fnhum.2015.00428

Takamura, Y., Imanishi, M., Osaka, M., Ohmatsu, S., Tominaga, T., Yamanaka, K., et al. (2016). Intentional gaze shift to neglected space: A compensatory strategy during recovery after unilateral spatial neglect. *Brain* 139, 2970–2982. doi: 10.1093/brain/aww226

Umarova, R. M., Beume, L., Reisert, M., Kaller, C. P., Kloppel, S., Mader, I., et al. (2017). Distinct white matter alterations following severe stroke: Longitudinal DTI study in neglect. *Neurology* 88, 1546–1555. doi: 10.1212/WNL.0000000000003843

Umarova, R. M., Nitschke, K., Kaller, C. P., Kloppel, S., Beume, L., Mader, I., et al. (2016). Predictors and signatures of recovery from neglect in acute stroke. *Ann. Neurol.* 79, 673–686. doi: 10.1002/ana.24614

Ye, L., Cao, L., Xie, H., Shan, G., Hu, J., Du, J., et al. (2020). Visual processing features in patients with visual spatial neglect recovering from right-hemispheric stroke. *Neurosci. Lett.* 714:134528. doi: 10.1016/j.neulet.2019.134528

Yin, D., Song, F., Xu, D., Peterson, B. S., Sun, L., Men, W., et al. (2012). Patterns in cortical connectivity for determining outcomes in hand function after subcortical stroke. *PLoS One* 7:e52727. doi: 10.1371/journal.pone.0052727

Zou, Q. H., Zhu, C. Z., Yang, Y., Zuo, X. N., Long, X. Y., Cao, Q. J., et al. (2008). An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: Fractional ALFF. *J. Neurosci. Methods* 172, 137–141. doi: 10.1016/j.jneumeth.2008.04.012

Frontiers in Neuroscience

TYPE Methods
PUBLISHED 18 August 2022
DOI 10.3389/fnins.2022.968928

EDITED BY
Xiaomao Fan,
South China Normal University, China

REVIEWED BY
Wenjun Tan,
Northeastern University, China
Ruxin Wang,
Shenzhen Institutes of Advanced Technology (CAS), China

*CORRESPONDENCE
Abdelkader Nasreddine Belkacem
belkacem@uaeu.ac.ae
Chunxiao Wan
Wcx2226@163.com

†These authors have contributed equally to this work

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

RECEIVED 14 June 2022
ACCEPTED 20 July 2022
PUBLISHED 18 August 2022

CITATION
Liang J, Song Y, Belkacem AN, Li F, Liu S, Chen X, Wang X, Wang Y and Wan C (2022) Prediction of balance function for stroke based on EEG and fNIRS features during ankle dorsiflexion. *Front. Neurosci.* 16:968928. doi: 10.3389/fnins.2022.968928

Prediction of balance function for stroke based on EEG and fNIRS features during ankle dorsiflexion

Jun Liang[1,2†], Yanxin Song[3†], Abdelkader Nasreddine Belkacem[4*], Fengmin Li[1], Shizhong Liu[1], Xiaona Chen[1], Xinrui Wang[1], Yueyun Wang[1] and Chunxiao Wan[1*]

[1]Department of Rehabilitation, Tianjin Medical University General Hospital, Tianjin, China, [2]Laboratory of Neural Engineering and Rehabilitation, Department of Biomedical Engineering, College of Precision Instruments and Optoelectronics Engineering, Tianjin University, Tianjin, China, [3]Tianjin Tianshi College, Tianjin, China, [4]Department of Computer and Network Engineering, College of Information Technology, United Arab Emirates University, Al Ain, United Arab Emirates

Balance rehabilitation is exceedingly crucial during stroke rehabilitation and is highly related to the stroke patients' secondary injuries (caused by falling). Stroke patients focus on walking ability rehabilitation during the early stage. Ankle dorsiflexion can activate the brain areas of stroke patients, similar to walking. The combination of electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS) was a new method, providing more beneficial information. We extracted the event-related desynchronization (ERD), oxygenated hemoglobin (HBO), and Phase Synchronization Index (PSI) features during ankle dorsiflexion from EEG and fNIRS. Moreover, we established a linear regression model to predict Berg Balance Scale (BBS) values and used an eightfold cross validation to test the model. The results showed that ERD, HBO, PSI, and age were critical biomarkers in predicting BBS. ERD and HBO during ankle dorsiflexion and age were promising biomarkers for stroke motor recovery.

KEYWORDS
brain-computer interface, EEG, fNIRS, stroke, balance rehabilitation

Introduction

Stroke is a disease affecting the arteries within the brain, resulting in motor impairment in about 80% of survivors (Langhorne et al., 2009). Among many stroke survivors, most patients were left with sequelae of motor dysfunction, and 30% of patients completely lost the ability to work and became highly disabled (Langhorne et al., 2009; Benjamin et al., 2017). Motor dysfunction causes patients to lose part of their living ability, rendering them unable to complete some daily living activities (Basteris et al., 2014). Therefore, motor recovery always focuses on stroke rehabilitation (Hatem et al., 2016). Balance recovery is essential to motor recovery, as the imbalance-leading falling substantially affects regular training and rehabilitation. In clinical practice, the Berg Balance Scale (BBS) is often used to evaluate the balance function of patients with cerebrovascular and brain injury (Sapmaz and Mujdeci, 2021). However, the scale's accuracy depends on the experience and subjective judgment of the physical therapists. A biomarker that can illustrate the balance recovery process is necessary to organize the rehabilitation strategy better and improve balance recovery. Developed imaging techniques have given valuable information for diagnostic and functional prognosis. Nevertheless, they may have limitations, such as the special requirements for patients and low temporal resolution (Mukherjee et al., 2008; Buchbinder, 2016). Therefore, more and more studies have concentrated on more convenient methods with electroencephalography (EEG) (Wu et al., 2016; Sebastian-Romagosa et al., 2020).

The EEG acquisition device is simple and portable and has a high temporal resolution. It is highly sensitive to detecting EEG activities and allows subjects to perform some complex limb movement tasks while observing them non-invasively and dynamically in real-time. The neurons' activity in the brain has been broadly used to monitor the stroke survivors' brain states (Cillessen et al., 1994; Foreman and Claassen, 2012; Xin et al., 2017). The EEG's beta band power patterns differed according to the location of the lesion (Park et al., 2016), and event-related desynchronization (ERD) magnitude correlated with residual motor function in the paretic arm (Bartur et al., 2019). However, one challenge of using EEG is its low spatial resolution problem, i.e., the ERD may be contaminated and weakened by the neural activities in the nearby areas. One alternative solution is to use functional near-infrared spectroscopy (fNIRS) as a supplement (Li et al., 2020). In a study using fNIRS to assess the correlation between cortical activation and external postural disturbances, the correlation became stronger with an increase in position-related oxygenated hemoglobin signal and an increase in balance function as measured by the BBS balance scale supplementary motor area (SMA) (Fujimoto et al., 2014). The fNIRS alone has been applied to assess the stroke's progressive brain plasticity (Delorme et al., 2019). It has also been used with EEG to estimate the effect of different training strategies (Wang et al., 2019). Therefore, combining fNIRS and EEG may give new sight to stroke rehabilitation assessment.

The stroke rehabilitation assessment with EEG or fNIRS was usually undertaken during resting tasks (Nicolo et al., 2015; Sebastian-Romagosa et al., 2020). However, motor recovery should be reflected better during motor or motor imagery tasks (Wang et al., 2019; Li et al., 2020) when the corresponding brain area is activated. Walking ability is an urgent need for stroke patients in the early stage. The assessment should be taken during walking to assess the walking ability of stroke patients precisely. Bipedal locomotion is a complex task requiring maintaining specific motion frequencies, balance and load-bearing, visual integration, and multi-joint coordination (Petersen et al., 2012). However, most stroke survivors during the early stage cannot walk, or they may fall off during walking.

Additionally, ankle dorsiflexion is critical for walking as it occurs throughout the swing phase and at the initiation of the stance phase of gait (Dobkin et al., 2004). How the stroke survivors complete the ankle dorsiflexion affects their walking ability. Therefore, ankle dorsiflexion may be a promising task for stroke rehabilitation assessment (Gennaro and De Bruin, 2020).

This paper aims to evaluate the combination of EEG and fNIRS features during ankle dorsiflexion in rehabilitation assessment. We collected data from stroke survivors during ankle dorsiflexion and built a linear regression model with age, ERD, and oxygenated hemoglobin (HBO) as the predictors and BBS as the response. Our results verified the feasibility of EEG and fNIRS combination in predicting stroke balance state.

Materials and methods

Participants and experiments

Eight participants (three females and five males; mean age: 53.5 ± 15.48 years old) with stroke participated. All participants suffered hemiplegia from the first unilateral stroke, resulting in lower limb function limitation without sensory function loss. They needed to understand written and oral instructions and be in a good mental state, with a mini-mental state examination (MMSE) score > 24. The healthy control group consisted of six age-matched adults. All participants gave their written informed consent prior to participation. The Ethical Committee of Tianjin Medical University General Hospital approved the study.

Before the experiment, the motor function of the participant was assessed by three physical therapists with the BBS scale for lower extremities. The mean of the scale values was recorded. The participants were seated in a chair during the experiment, with their feet naturally on the ground. They were asked to complete the paraplegic dorsiflexion according to the instructions on the screen as in **Figure 1**.

During the experiment, there were five sessions, each including ten trials. A single trial lasted for 11.5 s. Thus, it consisted of 1-s preparation, 2.5-s dynamic dorsiflexion, three-second static dorsiflexion maintenance, and five-second rest

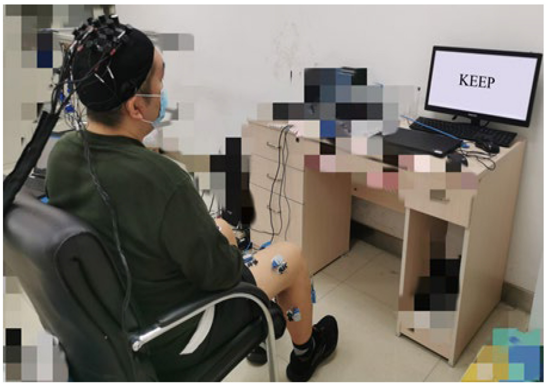


FIGURE 1
Experimental scenario.

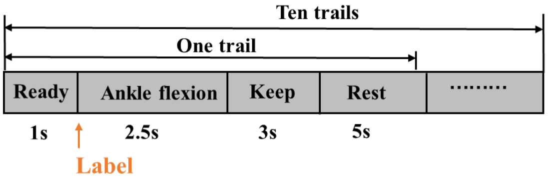


FIGURE 2
A session trail.

(Figure 2). A 5-min rest between successive sessions in case of muscle fatigue existed.

The participants were asked to relax their upper body, keep their upper body and head as still as possible, and avoid moving their heads, talking, swallowing, and blinking excessively during the dorsiflexion. The stimulation interface was completed based on E-Prime 3.0 software (Psychology Software Tools, Pittsburgh, PA, United States).

Data recording

EEG signals were recorded by the Neuroscan Greal EEG system (Neuroscan, Victoria, Australia). Twenty-eight-lead EEG signals were collected according to the 10/20 system, as in Figure 3A. The sampling rate of the EEG signal was 1,024 Hz, with the top of the head as a reference. The ground electrode was placed on the GND of the forehead.

Data processing

ERD

As ankle flexion was highly related to the activities of neurons under Cz, the collected raw EEG signals at Cz were pre-processed. Data pre-processing was performed using Matlab R2014b (Math Works, MA, United States) with the toolbox EEGLAB (Swartz Center for Computational Neuroscience[1]). EEG signals were filtered to 0.05–35 Hz, and the EEG was down-sampled to 256 Hz. Then, eye movements and excessive muscle activity components were identified by visual inspection and removed after independent component analysis (ICA). Data were intercepted from 1 s before to 5 s after the onset of ankle dorsiflexion (0 s), and −1 s to 0 s was the baseline.

The event-related power changes can be shown in the time-frequency domain by event-related spectral perturbation (ERSP), an excellent method to evaluate the time-frequency characteristics of event-related potentials. ERSP could provide information for event-related synchronization (ERS) and event-related desynchronization (ERD). It considers the average power spectrum changes of event-related potentials in a frequency band range. The ERSP for n-trial data was calculated as:

$$ERSP\left(f,t\right)=\frac{1}{n}\sum_{k=1}^{n}F_k\left(f,t\right)^2 \qquad (1)$$

where n indicated the total trial number, and $F_k\left(f,t\right)$ the spectral estimation at frequency f and time t for the kth trial.

To show ERD/ERS during the task, we calculated the baseline-normalized ERSP. For each time bin, the normalized ERSP across frequency f was obtained by

$$nERSP\left(f\right)=ERSP\left(f\right)-baseline\left(f\right) \qquad (2)$$

where $baseline\left(f\right)=\frac{1}{N}\sum_{t=-1s}^{0s}ERSP\left(f,t\right)$, N was the number of time bins from −1 s to 0 s. The ERD index, the predictor of the BBS scale value, was the mean of negative nERSP within a specific area (15 Hz∼23 Hz and 0 s∼1 s), determined by the average of nERSP in the time-frequency domain.

Phase synchronization index

The brain's function depends on the interaction between neurons in different regions or across brain regions (He et al., 2007). Recent studies have also demonstrated the efficacy of synchronized brain activity in evaluating neural networks and their relationship with various clinical conditions (Engel et al., 2013). Brain damage after a stroke can change brain function connections. Therefore, the multi-regional interactions in the brain network are valuable for understanding the pathophysiology and neurological dysfunction after a stroke (Du et al., 2018).

The phase synchronization index is a normalized parameter that measures the relationship between a pair of variables and effectively describes the integration between neurons. First, the pre-processed signal is filtered to the band of interest, and then

---

[1] http://sccn.ucsd.edu/eeglab/

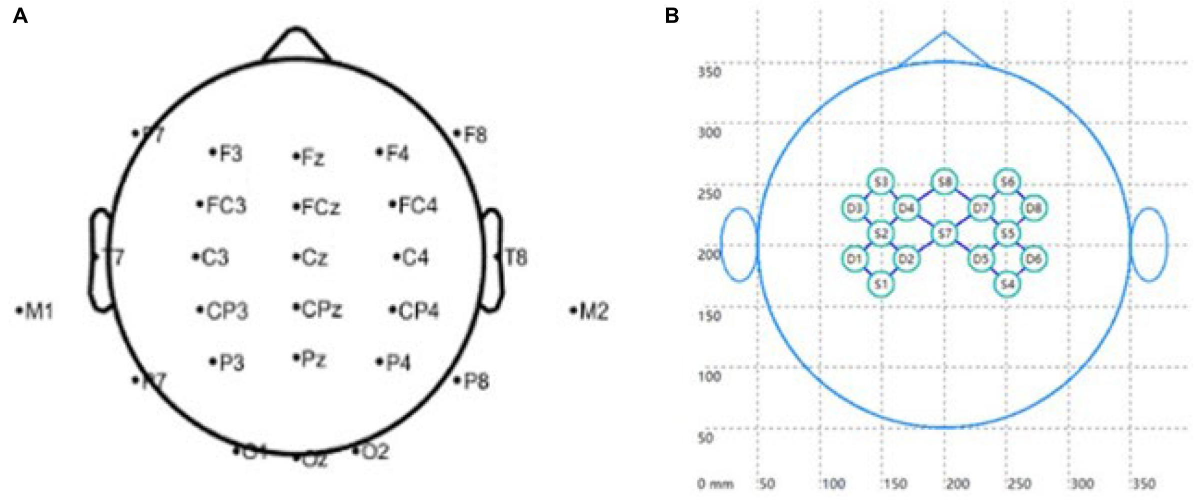


FIGURE 3
Signal acquisition. **(A)** EEG electrodes, and **(B)** fNIRS probe arrangement.

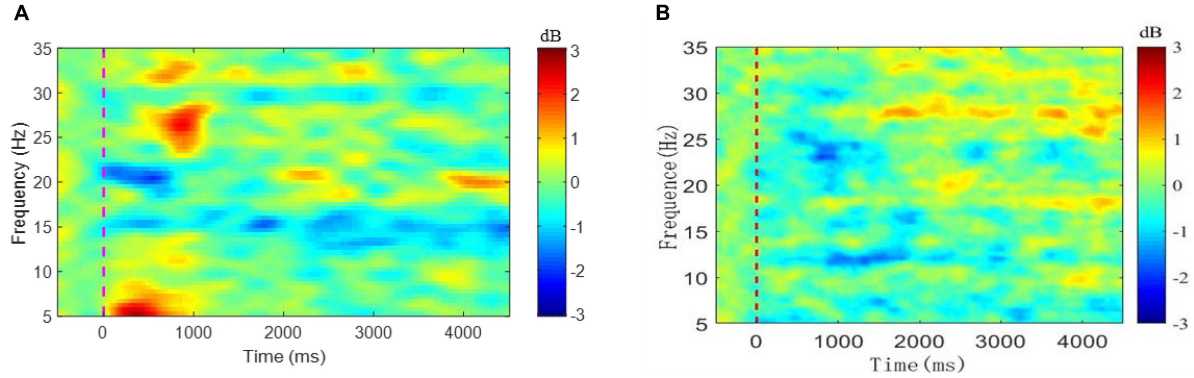


FIGURE 4
Mean of nERSP at Cz for **(A)** 8 patients and **(B)** 6 healthy control group.

the instantaneous phase of the filtered EEG signal is extracted using the Hilbert transform. The phase synchronization index is calculated as follows, quantifying the phase synchronization level of the signal:

$$PSI = \sqrt{\left\langle \cos \phi_{xy}^{H}(t) \right\rangle_t^2 + \left\langle \sin \phi_{xy}^{H}(t) \right\rangle_t^2} \quad (3)$$

which, $\phi_{xy}^{H}(t)$ is the instantaneous phase difference between the signals $x(t)$ and $y(t)$, and $\langle . \rangle$ represents the mean operation over some time. PSI is a real number between 0 and 1. When PSI = 1, the two signals are completely synchronized, and when PSI = 0, they are entirely out of sync.

A multi-channel oxygenation monitor (NirScan, Danyang Huichuang Medical Equipment Co., Ltd.) recorded the participants' hemodynamic responses over sensorimotor cortex areas. The distance between the source and the detector was 3 cm, and the sampling frequency was 20 Hz. The layout of the fNIRS acquisition cap is in **Figure 3B**, where S1-S8 represents fNIRS emission source probes and D1-D8 fNIRS detector probes. Thus, there were twenty-two fNIRS channels in total.

Oxygenated, deoxygenated, and total hemoglobin

The oxygenation monitor software calculated the oxygenated, deoxygenated, and total hemoglobin (HBO, HBR, and HBT). Then the data were 0.01 Hz~0.2 Hz filtered to exclude disturbance. The data within the area covered by S2-D4, S2-D2, S7-D2, S7-D4, S7-D7, S7-D5, S5-D7, and S5-D5 were segmented from 2 s before (baseline) to 10 s after ankle flexion. The task-related changes of HBO, HBR, and HBT according to the 2-s baseline were calculated and averaged for the above

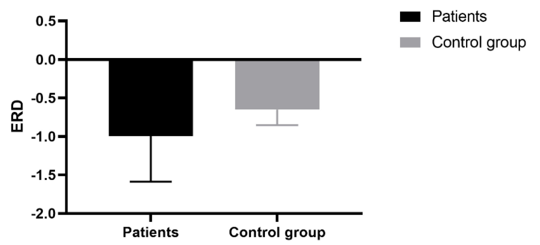


FIGURE 5
ERD index of patients and the control group.

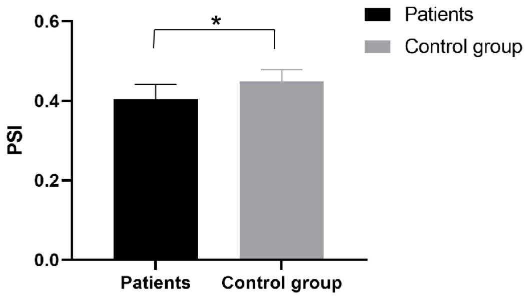


FIGURE 6
PSI index of patients and the control group. *Significant difference.

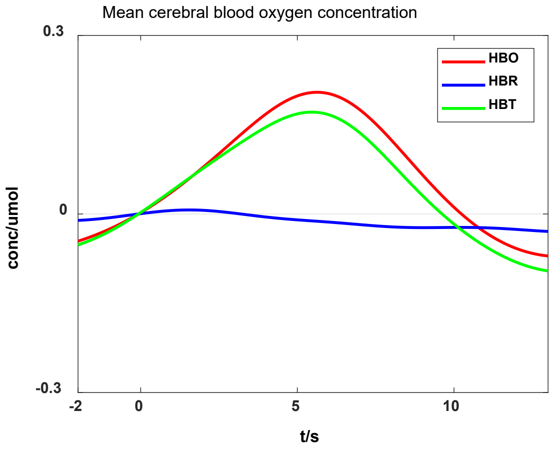


FIGURE 7
Cerebral blood oxygen concentration during ankle flexion.

fNIRS channels. The area between the average HBO curve and the horizontal axis predicted the BBS scale values.

Linear regression method

Linear regression analysis is a common modeling method; its fundamental principle is to use one or more independent variables to predict a dependent variable to establish a linear relationship. Assuming that the response variable Y is a linear function of one or more predictive variables (explanatory variables) X, multiple linear regression can be expressed as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + e \tag{4}$$

where Y is the target variable, β_i (i = 0, 1, 2. n) is the regression parameters, X_i = (i = 0, 1, 2. n) is the predictive variable of regression. The e is the error term.

After building the model, it is usually necessary to analyze whether the linear influence of the predictive variable X_i on the target variable Y is significant further. Next, we select the predictive variable with the most significant effect on the target variable Y and then determine the relative weight of each independent variable on the target variable.

This paper explored the correlation between BBS and multiple variables, including age, ERD, HBO and the square of HBO. A 4 × 8 variable existed for linear regression. The coefficient of determination, R^2, was used to evaluate the goodness of fit and statistical significance for each predictor with a pre-set alpha level of 0.05 for analysis. In order to further demonstrate and evaluate the effectiveness of the model, an eightfold cross validation of the model was done, with seven samples as the data for model establishment, and one sample for testing. The RMSE was calculated for the cross validation.

Results

ERD features

Figure 4A depicts the mean nERSP at Cz in a total of 400 tests of ankle flexion in eight stroke patients, while **Figure 4B** illustrates the mean nERSP at Cz in a total of 300 tests of ankle flexion in six healthy controls. Blue represents ERD and red ERS region. The ankle enters a dynamic buckling process within 2.5 s after 0 s and a static contraction after 2.5 s considering −1 s~0 s as the baseline. The figure reveals that the patient had an apparent ERD phenomenon at about 15 Hz (lasting the entire exercise cycle) and about 20 Hz (mainly during dynamic dorsiflexion within 1 s after the start of the movement and declining during the rest period). However, the evident ERD phenomenon of healthy subjects appeared at about 13 Hz and 25 Hz, and later than patients, and lasted shorter in the whole movement cycle. ERD phenomenon showed that cortical regions of the Cz electrode are convoluted in task planning and motion control during the task process, aligning with previous studies (Xu et al., 2018). Results verified that the paralyzed ankle flexion induced ERD within the beta band at Cz.

This study used the ERD index as the predictor of the BBS scale, the mean value of negative nERSP within a specific area (15 Hz~23 Hz and 0 s~1 s) determined by the mean nERSP in the time-frequency domain. In **Figure 5**, the calculated mean

TABLE 1 Age, ERD, HBO, and BBS of the patients.

Patient	Age	ERD	HBO	BBS
P1	60	−1.834	1.905	55
P2	74	−0.396	1.999	33
P3	49	−1.977	0.211	40
P4	64	−0.695	0.098	8
P5	33	−0.979	2.925	46
P6	32	−0.448	2.066	55
P7	67	−0.825	1.170	40
P8	49	−0.790	2.000	40
Avg.	54	−0.993	1.547	39.6

ERD of 8 patients was higher than that of the healthy controls. However, no significant difference between the two, possibly due to the limited sample size.

Phase synchronization index

We compared the phase synchronization index of EEG signals. We found significant differences ($p = 0.023$) in inter-PSI between patients and healthy subjects at the α band (**Figure 6**), a better understanding the differences between patients and healthy subjects during exercise. It suggested the stroke inactivation of the cortical network. The difference in PSI between hemispheres also indicated that the dynamic mode of synchronous activation between the two hemispheres deteriorated in functional communication due to the brain injury.

HBO features

Figure 7 shows that the mean HBO, HBR, and HBT values varied during ankle flexion: the HBO and HBT increased slowly after the ankle flexion onset (0 s) and decreased after the flexion completion (5.5 s). The fNIRS data differed consistently with ankle flexion, but the change lasted longer than the movement. HBO closely relates to the metabolism of local tissues. Therefore, the measured HBO can reflect the motor area activities.

Regression analysis

Table 1 lists the peak of the HBO curve for every patient. Then, this index should reflect the oxygen consumption of the specific brain area. It also depicts the ERD indexes at Cz of eight participants and the corresponding average values. We used the values in this table for the regression analysis.

The multiple linear regression model was established using the data in **Table 1**. **Table 2** lists the beta coefficient and p-values for all the model parameters. The adjusted R^2 was 0.840, indicating the model explained 84% of the variability in BBS. Correlation analysis revealed that no significant correlation between any two of the three predictors existed.

TABLE 2 Multiple linear regression model.

Independent variable	β (95% CL)	*p* value
Constant	30.6 (17.2, 44.0)	0.107
ERD	−12.4 (−16.5, −8.4)	0.054
HBO	37.3 (28.9, 45.8)	0.021
AGE	−0.5 (−0.7, −0.3)	0.075
HBO*HBO	−10.6 (−13.7, −7.4)	0.043

*Multiplication sign.

The regression model showed that the predictors ERD, AGE, and HBO could predict patients' balance function (**Figure 8**), and the prediction model was significant ($p = 0.043$).

The result of cross validation was shown in **Figure 9**. The RMSE of the predicted and actual BBS scores was 9.83 for the cross validation.

Discussion and conclusion

We focused on event-related desynchronization (ERD) as a motor command. ERD is a phenomenon where the α and/or the β band voltages decrease as the number of synchronized neural assemblies increases. Theα and β ERD occur before and during motor execution or motor imagery (Babiloni et al., 1999). Therefore, ERD can be interpreted as an electrophysiological correlate of activated cortical areas involved in processing sensory or cognitive information or the production of motor behavior (Pfurtscheller, 1992). An increased and/or more widespread ERD could result from the involvement of a more extensive neural network or more cell assemblies in information processing. Factors contributing to such an enhancement of the ERD include increased task complexity and more efficient task performance (Defebvre et al., 1996). This study covers dorsiflexion of the ankle joint as simple and easy for healthy subjects. However, they require more attention and energy for stroke patients due to motor dysfunction. It may also explain why ERD is more pronounced in patients than in healthy subjects. The ERD of healthy people during dynamic contraction of the ankle joint is more noticeable than that during static contraction, which may be due to activity in primary sensorimotor areas increasing in association with learning a new motor task and decreasing after the task has been learned (Zhuang et al., 1997).

In the fNIRS study, when subjects performed the ankle flexion task, the sensorimotor area of the cerebral cortex consumed oxygen and energy. At this point, the over-compensation mechanism of the brain blood supply system would flood the region with oxygen-rich blood, increasing HBO concentration and decreasing HBR. It further suggests

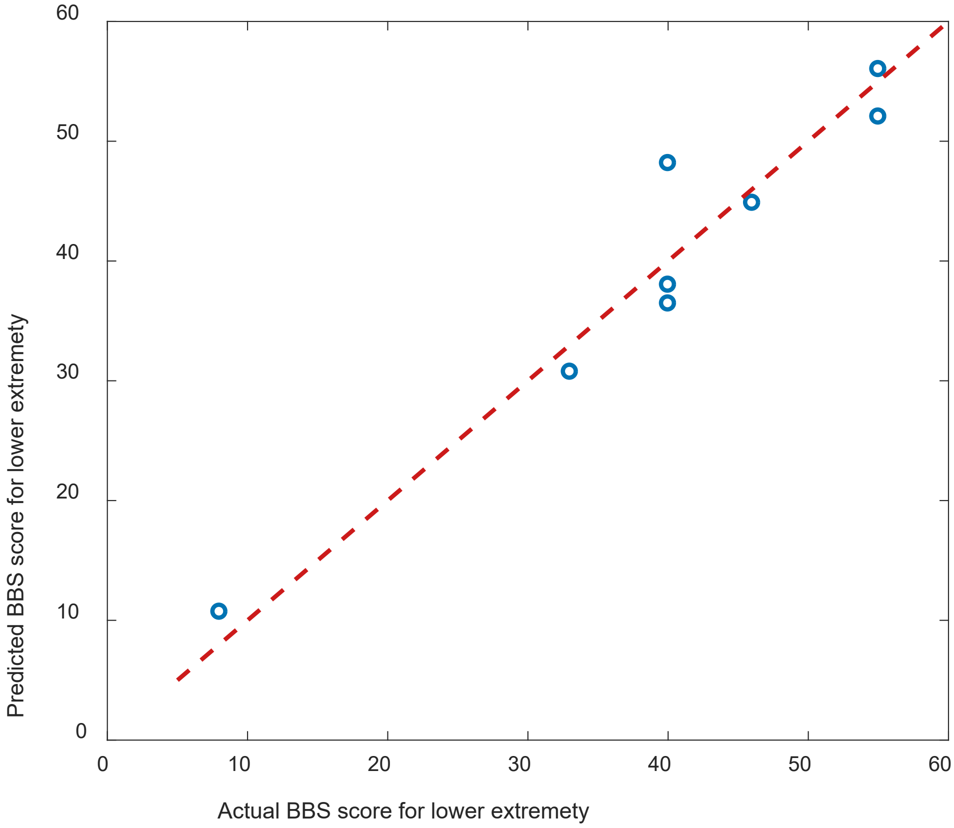


FIGURE 8
Scatterplots showing the relations between the actual and predicted BBS scales.

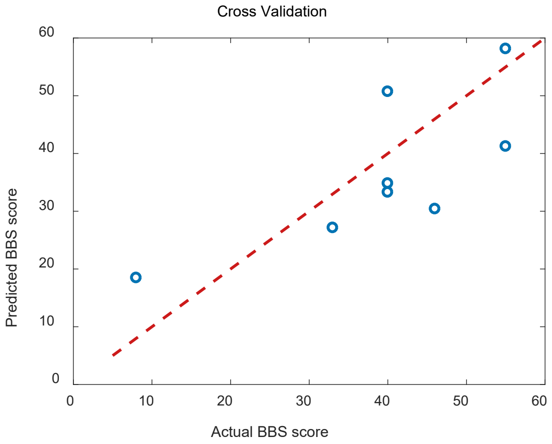


FIGURE 9
Scatterplots showing the relations between the actual and predicted BBS scales for cross validation.

that the experimental task activated this brain region. Previous studies have demonstrated that EEG and fNIRS-based brain-computer interface technology can enable a broader range of cortical activation in stroke patients, enhancing neuroplasticity (Kaiser et al., 2014). In this study, EEG features, including ERD, fNIRS, and HBO, can predict patients' balance function, providing a new idea for guiding the rehabilitation of stroke patients and evaluating and predicting patients' rehabilitation status. We considered the mean ERD and HBO values in this study. However, more features should be extracted from the data providing varied information. Future studies should address these features. Only eight stroke patients and six healthy control groups were recruited for our experiment, providing a relatively small number of subjects. The result of the regression analysis should be further validated after collecting more data from patients and healthy control.

To sum up, it was the first study considering both EEG and fNIRS features during ankle dorsiflexion as the biomarkers for stroke assessment. We extracted ERD and HBO features from eight stroke patients and established a linear regression model to predict BBS scale values. Age, ERD, and HBO during ankle dorsiflexion were promising biomarkers for stroke motor recovery. Further studies should include more participants with stroke and healthy controls to obtain a reliable relationship between these features and motor function state.

Data availability statement

The original contributions presented in this study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

Ethics statement

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

Author contributions

JL and FL completed the ethics files of this experiment. JL, YS, and SL recorded the original experiment data, analyzed the experiment data, and penned the manuscript. YW and AB wrote parts of the manuscript. CW designed the experiment and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was financially supported by the National Natural Science Foundation of China (82101448). AB was supported by the ASPIRE Young Investigator Award (AYIA 2020), funded by the ASPIRE under contracts 21T057-AYIA20-002, and by the United Arab Emirates University (Grant No. 31T130).

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

References

Babiloni, C., Carducci, F., Cincotti, F., Rossini, P. M., Neuper, C., Pfurtscheller, G., et al. (1999). Human movement-related potentials vs desynchronization of EEG alpha rhythm: A high-resolution EEG study. *Neuroimage* 10, 658–665. doi: 10.1006/nimg.1999.0504

Bartur, G., Pratt, H., and Soroker, N. (2019). Changes in mu and beta amplitude of the EEG during upper limb movement correlate with motor impairment and structural damage in subacute stroke. *Clin. Neurophysiol.*130, 1644–1651. doi: 10.1016/j.clinph.2019.06.008

Basteris, A., Nijenhuis, S. M., Stienen, A. H. A., Buurke, J. H., Prange, G. B., Amirabdollahian, F., et al. (2014). Training modalities in robot-mediated upper limb rehabilitation in stroke: A framework for classification based on a systematic review. *J. Neuroeng. Rehabil.* 11:111. doi: 10.1186/1743-0003-11-111

Benjamin, E. J., Blaha, M. J., Chiuve, S. E., Cushman, M., Das, S. R., Deo, R., et al. (2017). Heart disease and stroke statistics-2017 update a report from the American heart association. *Circulation* 135, E146–E603.

Buchbinder, B. R. (2016). Functional magnetic resonance imaging. *Handb. Clin. Neurol.* 135, 61–92.

Cillessen, J. P., Van Huffelen, A. C., Kappelle, L. J., Algra, A., and van Gijn, J. (1994). Electroencephalography improves the prediction of functional outcome in the acute stage of cerebral ischemia. *Stroke* 25, 1968–1972.

Defebvre, L., Bourriez, J. L., Destee, A., and Guieu, J. D. (1996). Movement related desynchronisation pattern preceding voluntary movement in untreated Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* 60, 307–312. doi: 10.1136/jnnp.60.3.307

Delorme, M., Vergotte, G., Perrey, S., Froger, J., and Laffont, I. (2019). Time course of sensorimotor cortex reorganization during upper extremity task accompanying motor recovery early after stroke: An fNIRS study. *Restor. Neurol. Neurosci.* 37, 207–218. doi: 10.3233/RNN-180877

Dobkin, B. H., Firestine, A., West, M., Saremi, K., and Woods, R. (2004). Ankle dorsiflexion as an fMRI paradigm to assay motor control for walking during rehabilitation. *Neuroimage* 23, 370–381. doi: 10.1016/j.neuroimage.2004.06.008

Du, J., Hu, J., Hu, J., Xu, Q., Zhang, Q., Liu, L., et al. (2018). Aberrances of cortex excitability and connectivity underlying motor deficit in acute stroke. *Neural Plast.* 2018:1318093. doi: 10.1155/2018/1318093

Engel, A. K., Gerloff, C., Hilgetag, C. C., and Nolte, G. (2013). Intrinsic coupling modes: Multiscale interactions in ongoing brain activity. *Neuron* 80, 867–886.

Foreman, B., and Claassen, J. (2012). Quantitative EEG for the detection of brain ischemia. *Crit. Care* 16:216.

Fujimoto, H., Mihara, M., Hattori, N., Hatakenaka, M., Kawano, T., Yagura, H., et al. (2014). Cortical changes underlying balance recovery in patients with hemiplegic stroke. *Neuroimage* 85, 547–554. doi: 10.1016/j.neuroimage.2013.05.014

Gennaro, F., and De Bruin, E. D. (2020). A pilot study assessing reliability and age-related differences in corticomuscular and intramuscular coherence in ankle dorsiflexors during walking. *Physiol. Rep.* 8:e14378. doi: 10.14814/phy2.14378

Hatem, S. M., Saussez, G., Della Faille, M., Prist, V., Zhang, X., Dispa, D., et al. (2016). Rehabilitation of motor function after stroke: A multiple systematic review focused on techniques to stimulate upper extremity recovery. *Front. Hum. Neurosci.* 10:442. doi: 10.3389/fnhum.2016.00442

He, B. J., Snyder, A. Z., Vincent, J. L., Epstein, A., Shulman, G. L., and Corbetta, M. (2007). Breakdown of functional connectivity in frontoparietal networks underlies behavioral deficits in spatial neglect. *Neuron* 53, 905–918.

Kaiser, V., Bauernfeind, G., Kreilinger, A., Kaufmann, T., Kübler, A., Neuper, C., et al. (2014). Cortical effects of user training in a motor imagery based brain-computer interface measured by fNIRS and EEG. *Neuroimage* 85, 432–444. doi: 10.1016/j.neuroimage.2013.04.097

Langhorne, P., Coupar, F., and Pollock, A. (2009). Motor recovery after stroke: A systematic review. *Lancet Neurol.* 8, 741–754.

Li, R., Li, S., Roh, J., Wang, C., and Zhang, Y. (2020). Multimodal neuroimaging using concurrent EEG/fNIRS for poststroke recovery assessment: An exploratory study. *Neurorehabil. Neural Repair* 34, 1099–1110. doi: 10.1177/1545968320969937

Mukherjee, P., Berman, J. I., Chung, S. W., Hess, C. P., and Henry, R. G. (2008). Diffusion tensor MR imaging and fiber tractography: Theoretic underpinnings. *Am. J. Neuroradiol.* 29, 632–641.

Nicolo, P., Rizk, S., Magnin, C., Pietro, M. D., Schnider, A., and Guggisberg, A. G. (2015). Coherent neural oscillations predict future motor and language improvement after stroke. *Brain* 138, 3048–3060. doi: 10.1093/brain/awv200

Park, W., Kwon, G. H., Kim, Y.-H., Lee, J. H., and Kim, L. (2016). EEG response varies with lesion location in patients with chronic stroke. *J. Neuroeng Rehabil.* 13:21. doi: 10.1186/s12984-016-0120-2

Petersen, T. H., Willerslev-Olsen, M., Conway, B. A., and Nielsen, J. B. (2012). The motor cortex drives the muscles during walking in human subjects. *J. Physiol.* 590, 2443–2452.

Pfurtscheller, G. (1992). Event-related synchronization (ERS): An electrophysiological correlate of cortical areas at rest. *Electroencephalogr. Clin. Neurophysiol.* 83, 62–69. doi: 10.1016/0013-4694(92)90133-3

Sapmaz, M., and Mujdeci, B. (2021). The effect of fear of falling on balance and dual task performance in the elderly. *Exp. Gerontol.* 147:111250.

Sebastian-Romagosa, M., Udina, E., Ortner, R., Dinarès-Ferran, J., Cho, W., Murovec, N., et al. (2020). EEG biomarkers related with the functional state of stroke patients. *Front. Neurosci.* 14:582. doi: 10.3389/fnins.2020.00582

Wang, Z., Zhou, Y., Chen, L., Gu, B., Yi, W., Liu, S., et al. (2019). BCI monitor enhances electroencephalographic and cerebral hemodynamic activations during motor training. *IEEE Trans. Neural. Syst. Rehabil. Eng.* 27, 780–787.

Wu, J., Srinivasan, R., Quinlan, E. B., Solodkin, A., Small, S. L., and Cramer, S. C. (2016). Utility of EEG measures of brain function in patients with acute stroke. *J. Neurophysiol.* 115, 2399–2405.

Xin, X., Chang, J., Gao, Y., and Shi, Y. (2017). Correlation between the revised brain symmetry index, an eeg feature index, and short-term prognosis in acute ischemic stroke. *J. Clin. Neurophysiol.* 34, 162–167. doi: 10.1097/WNP.0000000000000341

Xu, R., Wang, Y., Wang, K., Zhang, S., He, C., and Ming, D. (2018). Increased corticomuscular coherence and brain activation immediately after short-term neuromuscular electrical stimulation. *Front. Neurol.* 9:886. doi: 10.3389/fneur.2018.00886

Zhuang, P., Toro, C., Grafman, J., Manganotti, P., Leocani, L., and Hallett, M. (1997). Event-related desynchronization (ERD) in the alpha frequency during development of implicit and explicit learning. *Electroencephalogr. Clin. Neurophysiol.* 102, 374–381. doi: 10.1016/s0013-4694(96)96030-7

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 01 September 2022
DOI 10.3389/fnins.2022.971829

OPEN ACCESS

EDITED BY
Xiaomao Fan,
South China Normal University, China

REVIEWED BY
Xianhui Chen,
Tandon School of Engineering, New York University, United States
Tian Jian Luo,
Fujian Normal University, China

*CORRESPONDENCE
Xiaoling Wang
xlwang@cs.ecnu.edu.cn

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

RECEIVED 17 June 2022
ACCEPTED 04 August 2022
PUBLISHED 01 September 2022

CITATION
Hu M, Chen J, Jiang S, Ji W, Mei S, Chen L and Wang X (2022) E2SGAN: EEG-to-SEEG translation with generative adversarial networks. *Front. Neurosci.* 16:971829. doi: 10.3389/fnins.2022.971829

E2SGAN: EEG-to-SEEG translation with generative adversarial networks

Mengqi Hu[1], Jin Chen[2], Shize Jiang[3], Wendi Ji[1], Shuhao Mei[3], Liang Chen[3] and Xiaoling Wang[1]*

[1]School of Computer Science and Technology, East China Normal University, Shanghai, China,
[2]Institute for Biomedical Informatics, University of Kentucky Lexington, Lexington, KY, United States,
[3]Department of Neurosurgery, Huashan Hospital of Fudan University, Shanghai, China

High-quality brain signal data recorded by Stereoelectroencephalography (SEEG) electrodes provide clinicians with clear guidance for presurgical assessments for epilepsy surgeries. SEEG, however, is limited to selected patients with epilepsy due to its invasive procedure. In this work, a brain signal synthesis framework is presented to synthesize SEEG signals from non-invasive EEG signals. First, a strategy to determine the matching relation between EEG and SEEG channels is presented by considering both signal correlation and spatial distance. Second, the EEG-to-SEEG generative adversarial network (E2SGAN) is proposed to precisely synthesize SEEG data from the simultaneous EEG data. Although the widely adopted magnitude spectra has proved to be informative in EEG tasks, it leaves much to be desired in the setting of signal synthesis. To this end, instantaneous frequency spectra is introduced to further represent the alignment of the signal. Correlative spectral attention (CSA) is proposed to enhance the discriminator of E2SGAN by capturing the correlation between each pair of EEG and SEEG frequencies. The weighted patch prediction (WPP) technique is devised to ensure robust temporal results. Comparison experiments on real-patient data demonstrate that E2SGAN outperforms baseline methods in both temporal and frequency domains. The perturbation experiment reveals that the synthesized results have the potential to capture abnormal discharges in epileptic patients before seizures.

KEYWORDS

EEG-SEEG mapping, GANs, epilepsy, signal synthesis, stereoelectroencephalography, deep learning

1. Introduction

Stereoelectroencephalography (SEEG) is an intracranial recording that can pinpoint the areas of the brain where seizures occur (Chabardes et al., 2018). SEEG signals are acquired by depth electrodes implanted into the brain (Li et al., 2018). Its high spatial and temporal resolution enable the recording of high-amplitude and high-frequency intracranial discharges that are sometimes difficult to observe on scalp electroencephalogram (EEG) (Ramantani et al., 2016). Despite its enormous benefits, SEEG is yet to be a panacea. To implant the electrodes, an invasive surgical procedure is required to make 10–20 small incisions in the scalp and skull. SEEG is only reserved for selected epilepsy patients due to the potential risk of infection (Cossu et al., 2005).

EEG, on the other hand, is an electrophysiological recording of electrical activity on the scalp by placing electrodes in fixed positions (Henry, 2006). EEG is non-invasive, relatively safe, inexpensive, functionally fast, and has been widely used to observe the spontaneous electrical activity of the brain. The electromagnetic fields recorded by EEG represent the linear summation of collective source activity (Plummer et al., 2008; He et al., 2019). Nevertheless, its relatively low signal-to-noise ratio, due to the attenuation by the layers lying around the brain, hinders the use of EEG for accurate epilepsy diagnoses.

In order to obtain intracranial signal recordings at low risk, a feasible solution is to recover an intracranial signal from a low-cost non-invasive signal. Cao et al. (2022) introduce the concept of virtual intracranial EEG (ViEEG) to reconstruct electrocorticography (ECoG) from magnetoencephalographic imaging (MEG). Dynamical network models are then applied to ViEEG to probe the underlying mechanisms of complex neural dynamics. Compared with ECoG, SEEG acts as an intracranial signal in the same way with richer spatial resolution, and the reconstruction is more challenging as well. As a collection of intracranial signals in the scalp, EEG is thought to be closely related to SEEG (Ramantani et al., 2016). Inspired by previous work and supported by the existing medical background, we propose a solution to synthesize intracranial SEEG from non-invasive EEG to face the above challenge and define this challenge as EEG-to-SEEG translation which is shown in Figure 1. The synthesized SEEG should retain the key features of real SEEG. In particular, the key features should carry clinical implications that can be regarded as plausible explanations of specific intracranial electrophysiological activity such as abnormal epileptic discharges. By indicating under what conditions key features are captured, clinicians can use the synthesized results in a targeted manner when assessing the need for SEEG implantation and then pinpointing the location for electrode implantation.

In recent years, the thriving of deep learning drives the development of various fields including EEG analysis, providing us with a new and feasible way of thinking. Antoniades et al. (2018) made an attempt to devise an asymmetric auto-encoder stacked with multi-layer neurons to map the temporal sequence of EEG to SEEG and outperform the previous linear methods such as least-squares regression (Kaur et al., 2014) and coupled dictionary learning (Spyrou and Sanei, 2016). The stacked architecture enhances the model's capacity but the simple auto-encoder architecture is still not powerful enough to achieve the desired result. Their attempt demonstrated the possibility of synthesizing SEEG from an input EEG and helped us recognize the necessity of a more powerful feature extractor and a sophisticated generation architecture. More recently, generative adversarial networks (GANs) (Goodfellow et al., 2014) have become a milestone in data generation and attracted our interest. GANs are basically composed of a generator network and a discriminator network. The process of adversarial training forces the generator to synthesize results with more details. As a result, the discriminator can hardly distinguish the real and generated ones. The ability of GANs to fit input and output distributions makes it outstanding in heterogeneous data synthesis (Jiao et al., 2019; Selim et al., 2020). The excellence of GANs soon inspired researchers to leverage this architecture as a new fashion to generate EEG signals (Hartmann et al., 2018; Luo T-j. et al., 2020; Yao et al., 2020). Furthermore, an improvement in GAN by imposing a condition on the input has achieved great success in image-to-image translation (Isola et al., 2017), which makes it possible to transfer the style or texture of the input to the output image. They utilized PatchGAN as a new paradigm of discriminator in order to restrict GANs to only model high-frequency structures. Their work enlightened us to leverage conditional GANs (cGANs) (Mirza and Osindero, 2014) together with PatchGAN-based paradigm to transfer an EEG segment to the SEEG segment.

Although the above methods are mature and proved to be effective, most of them only consider temporal representation, neglecting the informative features hidden behind. Numerous studies have demonstrated that time-frequency representations obtained from Short-Time Fourier Transform (STFT) (Li et al., 2021) or Morlet wavelet convolutions provide richer information and help give better predictions (Yao et al., 2018; Wang et al., 2020). Clinically, different EEG bands have discriminative implications (Tatum, 2014) and serve as beneficial features in many tasks (Yao et al., 2018; Wang et al., 2020). Therefore, in the context of EEG-to-SEEG translation, it is necessary to explore and exploit the correlation between two signals from a frequency domain perspective. Another discovery that has intrigued us is that the partial derivative of the unwrapped phase with respect to time, commonly referred to as the instantaneous frequency (IF), has great potential in the synthesis of phase spectra (Engel et al., 2019; Marafioti et al., 2019). Better phase spectrum synthesis ensures more coherent temporal results, which is one of the aims of this work.

Furthermore, SEEG electrodes outnumber EEG electrodes in most cases, and the placement of SEEG electrodes varies from patient to patient. Therefore, for the set-wise translation, it is inevitable to determine the matching relationship between the EEG and the SEEG set. This requires us to develop a strategy to select reasonable pairs from a large number of candidate fragments. The selected pairs are expected to contain clinically meaningful features. Subject to the complexity of EEG data, we were unable to accurately capture key signals such as potential pre-seizure micro-abnormal discharges. However, based on the good synchronization property of EEG and SEEG, it can be assumed that this key signal must be hidden in pairs with a strong correlation. In general, this correlation is affected by physical distance, and the strength can be measured by calculations based on power spectral density.

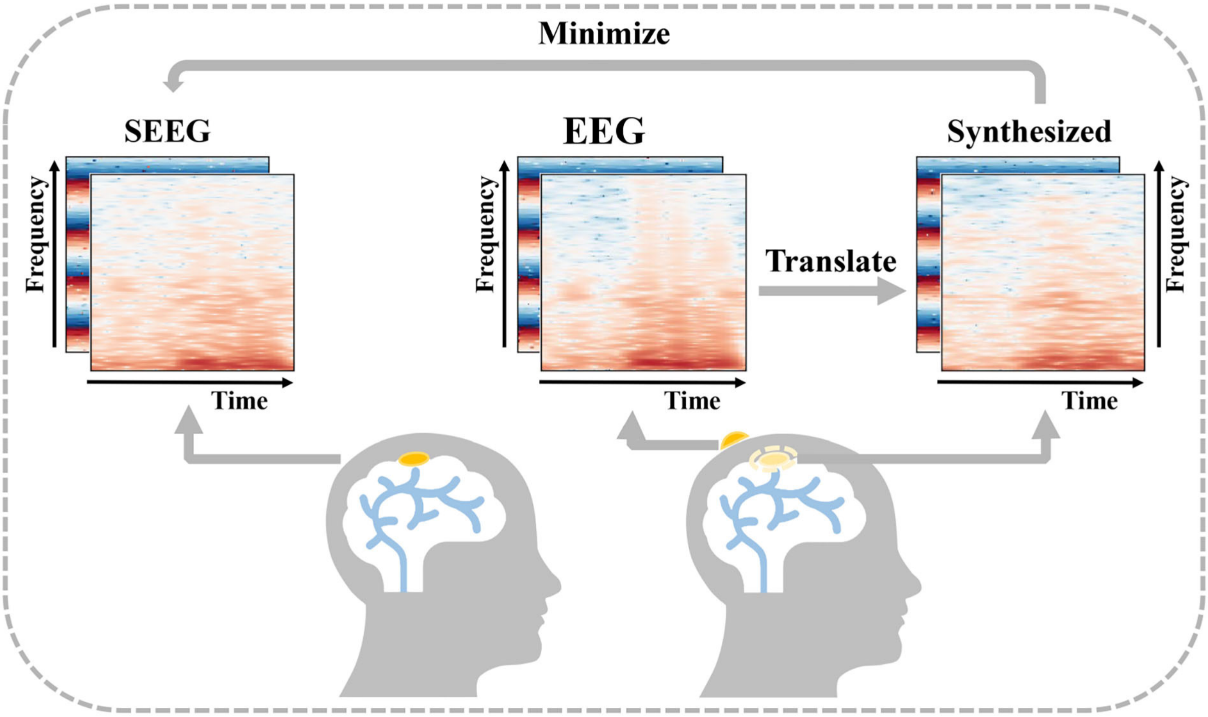


FIGURE 1
Description of EEG-to-Stereoelectroencephalography (SEEG) translation. EEG and SEEG are represented with a concatenated matrix of frequency and IF spectra. The difference between the real invasive SEEG and that synthesized from simultaneous non-invasive EEG is minimized.

In this work, by leveraging both the temporal and frequency characteristics of brain signals, an EEG-SEEG matching strategy is designed to construct an aligned dataset and an EEG-to-SEEG generative adversarial network called E2SGAN is proposed.

First, the EEG-SEEG matching strategy explores the nonlinear correlation between EEG and SEEG by observing how signal similarity varies with spatial distance. An aligned dataset is constructed with the selected pairs in the form of the time-frequency representation obtained from the STFT transform.

Second, the E2SGAN trained on the aligned dataset is proposed to convert the input EEG to the corresponding SEEG. The E2SGAN architecture takes full advantage of the time-frequency features and learns how to synthesize the magnitude and IF spectra accurately. The generator is built with residual blocks connecting a CNN-based encoder-decoder structure, and the discriminator is designed according to the patch-based paradigm. Two auxiliary modules, called correlative spectral attention (CSA) and weighted patch prediction (WPP), are devised to enhance the discriminator's ability. CSA captures the correlation between different combinations of EEG and SEEG frequencies and prevents the discriminator from making judgments based solely on the geometry of the spectra. WPP is a technique that eliminates potential mode collapse that occurs with each frequency to ensure more robust temporal results.

Extensive comparison experiments have shown that the proposed framework is able to outperform the baseline methods. The perturbation experiment reveals that the synthesized results have the potential to capture abnormal discharges in epileptic patients before seizures.

The main contributions of this work are as follows:

- We propose E2SGAN, a practical deep-learning algorithm to address the EEG-to-SEEG translation. CSA and WPP are devised to capture the correlation between EEG and SEEG spectra and ensure robust temporal results.
- We pioneer the introduction of both magnitude and IF spectra as a time-frequency representation in a brain signal conversion setup.
- We develop an EEG-SEEG matching strategy to determine the matching relation between EEG and SEEG sets. The strategy explores the nonlinear correlation of signal similarity with respect to spatial distance.
- Evaluation results on extensive real-patient-based experiments demonstrate the excellent performance of the proposed framework in both temporal and frequency domains. A further perturbation experiment reveals the

potential of the synthesized results to capture abnormal epileptic discharges.

2. Materials and methods

2.1. Framework overview for SEEG synthesis

A two-fold pipeline is designed for SEEG synthesis as depicted in Figure 2, (1) to prepare aligned training data by matching EEG and SEEG segments (see Section 2.2) and (2) to translate EEG to SEEG (see Section 2.3 to 2.3.6).

The first stage does a preliminary job of organizing preprocessed EEG and SEEG signals in pairs to form a temporally aligned dataset. Raw EEG and SEEG signals are filtered and segmented using a synchronized sliding window. Then, segments from two sources are matched into pairs with the proposed matching strategy (see Section 2.2) based on the correlation between signal similarity and physical distance. The paired segments are processed with STFT to obtain magnitude spectra and IF spectra. The aligned dataset is constructed with the processed pairs.

At the second stage, E2SGAN is trained on the aligned dataset. The generator transforms the input EEG spectra to the target SEEG spectra while the discriminator makes an effort to distinguish real and fake targets conditioned by the input EEG. CSA (see Section 2.3.4) and WPP (see Section 2.3.5) are two auxiliary modules devised to give a further boost to the discriminator. Specifically, CSA is a mutual attention sub-network that captures the correlation between EEG and SEEG frequencies. The captured correlation can be considered as an extra supervision signal which ensures the correct frequency correlation of the synthesized target with respect to the input. WPP is a customized technique to disturb the monotonous distribution of patch prediction. The variation within patch-based prediction alleviates the mode collapse caused by low variance in IF spectra. The optimization of the whole network is described in Section 2.3.6. In the end, the synthesized SEEG spectra are transformed back to temporal representation *via* inverse STFT. The implementation of the proposed framework is available at https://anonymous.4open.science/r/E2SGAN-180B/.

2.2. EEG-SEEG matching strategy

As is discussed in Section 1, a matching strategy aims to address two challenges. First, it has to make a compromise on the difficulty of directly translating the whole set of EEG to SEEG. Second, it should single out the potential pairs carrying implicit clinical features, which can be measured by the strength of correlation within a pair. Therefore, we settle for the second best to focus on one-to-one mapping within pairwise EEG and SEEG channels. Based on this setting, the set-wise translation is decomposed to sub-tasks where the pairwise translation will be performed. Such a strategy has to guarantee the existence of a correlation between EEG and SEEG segments within a pair. Specifically, a qualified solution should obey the following procedure:

1. Map the set of EEG channels to the set of SEEG channels *via* any form of bipartite graph matching to obtain a sea of candidate EEG-SEEG pairs
2. Search for the optimal pairs while ensuring the correlation
3. Generalize the solution so as to be applied to upcoming subjects

For this strategy, such correlation is defined as signal similarity based on a given similarity metric. The correlation does not vary with distance in a linear trend as is habitually deemed, but rather a non-linear fashion. To demonstrate this counterintuitive relation, the Hellinger Distance (Chen et al., 2020) from two SEEG channels and the target EEG channel are compared as shown in Figure 3. The leftmost column gives an example of what a matching strategy is expected to solve. When matching the EEG *Cz* (yellow-dotted at the top of Figure 3A), the strategy is making a decision between the choice of SEEG *A14* and *H14* (red-/green-circled at the bottom of Figure 3A). Despite the closer physical distance of the SEEG channel *A14* to EEG *Cz* as is shown in the topological map (top of Figure 3B) and the 3-D location (bottom of Figure 3B), the farther HD is observed in *A14* rather than *H14* in Figure 3C. Hence, *H14* is considered to be an appropriate match to *Cz*. Potential reasons can be the influence of brain geometry or brain functional connectivity on signal propagation (Frauscher et al., 2018). To guarantee a meaningful pairwise translation, a matching strategy has to take into account the complex correlation within a pair.

Here, one of the implementations following the proposed procedure is presented in Algorithm 1 which deals with the complex correlation in a simple but efficient way. Regarding the EEG and SEEG channels as two non-overlapped sets, the problem is initiated as bipartite graph matching. To fulfill Procedure 1, all the SEEG channels are first matched to a given EEG channel to obtain candidate EEG-SEEG pairs. In Algorithm 1, Euclidean distance is adopted as $d(\cdot)$ to sort the pairs. Considering the non-linear trend, the set of C is divided into subsets, leaving each subset corresponding with a distance interval *itv*.

For each subset, linear regression is used as f to explore the correlation between physical distance and signal similarity. Here, Hellinger Distance is adopted as $s(\cdot)$ and the first-order derivative ∇f is calculated, of which the positive value indicates that the similarity between the EEG channel and SEEG channels decreases as the physical distance becomes farther. The

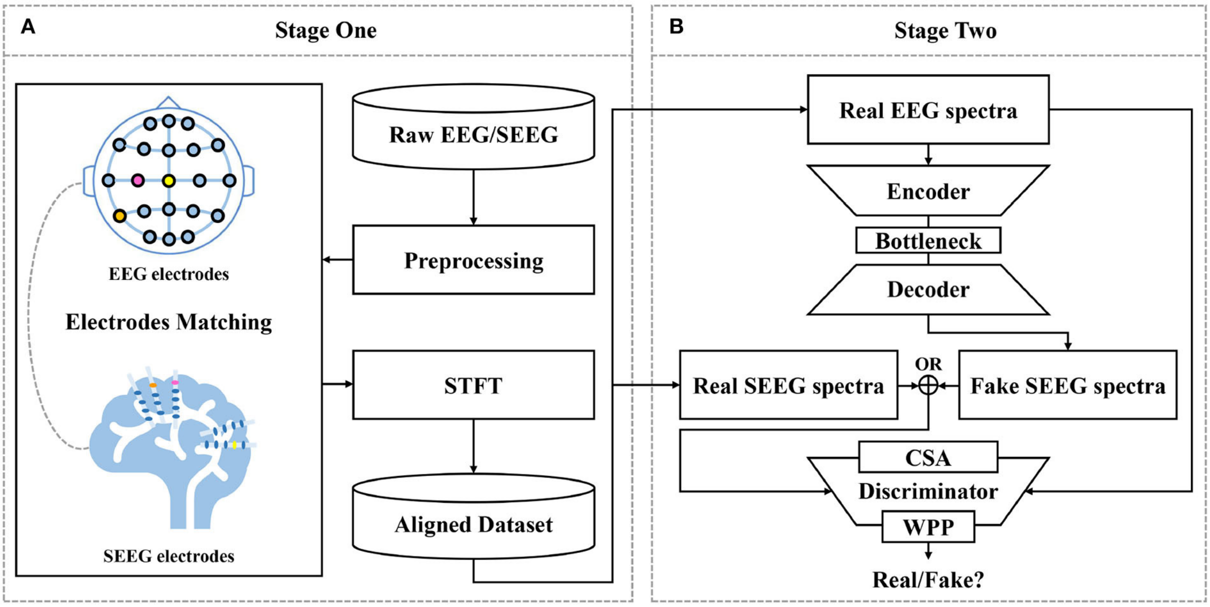


FIGURE 2
Proposed framework for EEG-to-SEEG translation. **(A)** Raw EEG and SEEG signals are filtered and segmented using a synchronized sliding window. Then, segments from two sources are matched into pairs with the proposed strategy. Short-Time Fourier Transform (STFT) is performed to obtain magnitude and IF spectra. The aligned dataset is constructed with the processed pairs. **(B)** EEG-to-SEEG generative adversarial network (E2SGAN) is trained on the aligned dataset to synthesize SEEG from simultaneous EEG. Correlative Spectral Attention (CSA) and Weighted Patch Prediction (WPP) are devised to give a further boost to the discriminator.

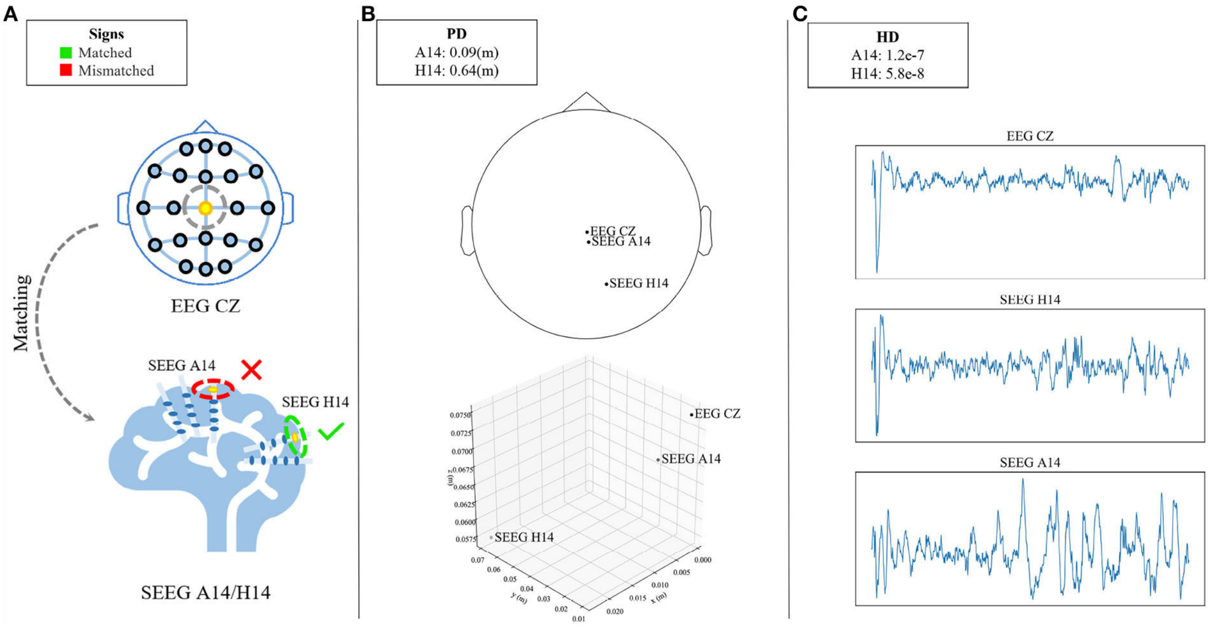


FIGURE 3
Demonstration of the counterintuitive relation between signal correlation and physical distance. **(A)** Illustrates that the matching strategy needs to match the EEG channel (*Cz*, yellow-doted at the top) to one of the candidate SEEG channels (*A14*/*H14*, red-/green-circled at the bottom). Despite the closer physical distance of SEEG *A14* to EEG *Cz* which is shown in the topological map (top of **B**) and the 3-D location (bottom of **B**), the farther HD is observed in *A14* rather than *H14* **(C)**. *H14* is considered to be an appropriate match to *Cz*.

```
Input: EEG channel $o$ and SEEG channels $C = \{c_1, c_2, ..., c_N\}$
Function: physical distance metric $d(\cdot)$, signal
similarity metric $s(\cdot)$, regression function $f$, and
correlation operator $\nabla$
Output: distance interval $itv^*$

1: Sort elements $c_i \in C$ by $d(o, c_i)$ in ascending order
2: Split $C$ into subsets so that each subset $C_{sub}$
   corresponds with a distance interval $itv$
3: for $C_{sub}$ of subsets do
4:     Obtain the correlation $\nabla f$ where $f: d(o, c_j) \rightarrow s(o, c_j)$,
       $c_j \in C_{sub}$
5: end for
6: Repeat 1 to 5 on all segments and subjects
7: return interval $itv^*$ where the most frequent
   correlation appears
```

Algorithm 1. EEG-SEEG Matching Strategy

proportion of positive ∇f is counted on each subset. The steps mentioned above are carried out on all segments and subjects. After that, the interval itv^* where the most positive derivatives appear, ranging from a to b mm, i.e., $itv^* = [a, b]$, is chosen. SEEG channel approximately a mm away from the investigated EEG is considered a match.

Within itv^*, the similarity between EEG and SEEG segments prominently declines as the physical distance increases. Latent correlations between SEEG and EEG signals have been shown within this interval, even not necessarily stronger than the others, thus satisfying Procedure 2. Other off-the-shelf methods for finding optimal EEG-SEEG similarity can be directly applied to the proposed strategy by replacing the metric $s(\cdot)$ in Algorithm 1.

In the following experiments, EEG *Cz* is chosen to obtain the output interval itv^*, and the conclusion is extended to all other EEG channels as is required by Procedure 3.

2.3. E2SGAN method

In the second stage, we center on the EEG to SEEG translation. In this section, some rudiments of instantaneous frequency and conditional GAN are provided as preliminary. Then, the details about the generative model are provided, including two auxiliary modules and the objectives.

2.3.1. Brain signal representation with instantaneous frequency

To better characterize brain signals, we employ STFT, which transforms the original signal into the frequency domain for the generation task. Different from the previous classification tasks, in addition to considering the energy distribution of the brain signal in the frequency domain, the phase distribution of the signal needs to be taken into consideration. Otherwise, the time domain representation of the signal cannot be restored. In practice, frame-based techniques used in signal process/generation such as deconvolutions will cause the initial phase of the segment intercepted by the frame to change over time when the stride of frames does not equal the signal's periodicity. This phenomenon renders the generation of phase spectra a tricky task as covering all the necessary frequencies and all possible phase alignments to preserve the phase coherence is impractical (Engel et al., 2019).

The partial derivative of the unwrapped phase with respect to time, commonly referred to as instantaneous frequency (IF), is a time varying measure of the true signal oscillation. For example, given a function of phase with respect to time

$$\varphi(t) = \omega t + \theta \tag{1}$$

where ω is the frequency and θ is the initial phase or phase offset. The instantaneous frequency is defined as

$$\omega(t) = \frac{d\varphi(t)}{dt} \tag{2}$$

where $\varphi(t)$ has to be in unwrapped form (Sejdic et al., 2008). In this case, a time-independent constant θ is derived. It has been proved that the instantaneous frequency of phase is a more promising modeling target than the phase itself when generating signals or phase spectra (Marafioti et al., 2019). The assumption of instantaneous frequency alleviates this risk since it remains constant on each frequency and is feasible to be learned by neural networks.

2.3.2. Conditional GANs for EEG-to-SEEG translation

The Generative Adversarial Network proposed by Goodfellow et al. (2014) has been proved to have strong data generation ability, which benefits from its unique network architecture: a pair of generator network and discriminator network competing with each other to generate indistinguishable data. The generator network is usually composed of an encoder-decoder structure, which takes noise sampled from a known prior distribution as input and aims to fit the target data distribution as precisely as possible. The discriminator network focuses on determining whether the input data comes from the target data distribution or is forged by the generator. Analogous to the game process, the two networks optimize their own parameters through the feedback given by each other and finally produce an output that is indistinguishable from real and fake.

To control the modes of the data being generated, Mirza and Osindero (2014) proposed the conditional version of GANs.

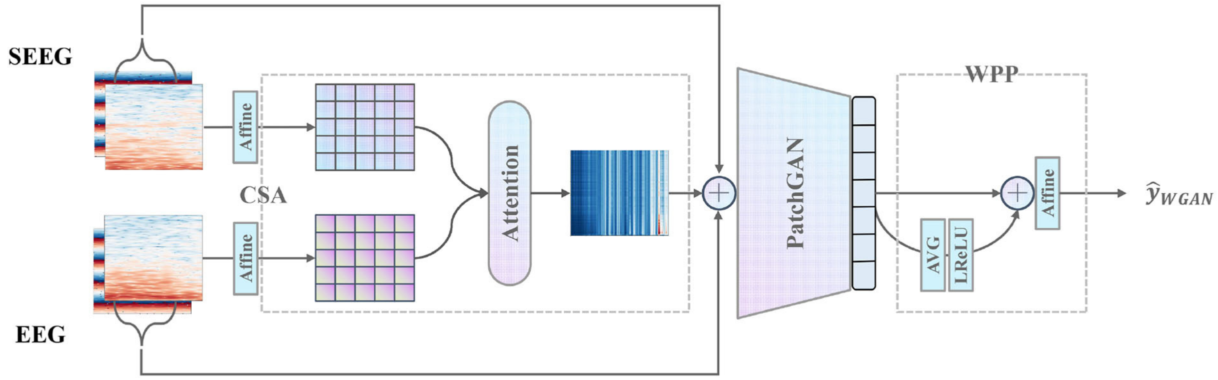


FIGURE 4
Proposed discriminator with the pre-positioned CSA module capturing the frequency correlation of EEG and SEEG, and the post-positioned WPP module ensuring robust temporal results.

Given a generator network G and a discriminator network D, the optimization objective of cGANs can be formulated as follows:

$$\mathcal{L}_{cGAN} = \mathbb{E}_{\boldsymbol{x}\sim p_{data}(\boldsymbol{x})}[\log D(\boldsymbol{x}|\boldsymbol{y})] + \mathbb{E}_{\boldsymbol{z}\sim p_z(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z}|\boldsymbol{y})))] \tag{3}$$

where $\boldsymbol{x}$ is a sample from the target distribution, $\boldsymbol{y}$ is the condition, and $\boldsymbol{z}$ is sampled from a prior distribution. The core idea of cGANs is to concatenate an extra condition to the inputs of both G and D, imposing the networks to determine whether the generated data matches the given condition. If $\boldsymbol{y}$ also conforms to a known distribution, we are able to realize the conversion between the two different data distributions. To ensure a one-to-one mapping, the addition of noise needs to be removed.

The basic idea of cGANs is adopted to achieve the goal of translating EEG to SEEG. Overall, the generator G establishes a mapping from an EEG segment e to an SEEG segment s where e, $s \in \mathbb{R}^{2\times m\times n}$. The first dimension represents the concatenated magnitude and IF spectra. m and n are the numbers of frequencies and time steps after the STFT operation. Subsequently, D takes the (e, s) pair as input where s is either a real SEEG segment or generated and outputs a scalar to indicate the difference between the distributions of real and fake pairs. PatchGAN (Isola et al., 2017) is adopted as D, which is a fully convolutional neural network that penalizes structure at the scale of patches based on the Markov chain assumption. Noise added to the input is omitted since the task is a determinate one-to-one mapping.

Furthermore, CSA and WPP are devised to give a further boost to the discriminator as is shown in Figure 4. CSA is a mutual attention sub-network that captures the correlation between each combination of EEG and SEEG frequencies. The captured correlation can be considered an extra supervision signal which ensures the correct frequency correlation of the synthesized target with respect to the input. WPP is a customized technique to disturb the monotonous distribution of patch predictions. The variation within patch-based prediction alleviates the mode collapse caused by low variance in IF spectra and ensures robust temporal results.

2.3.3. Architecture of generator and discriminator networks

The architectures of the generator and discriminator networks are shown in Table 1. The generator is basically a three-layer CNN autoencoder with two residual blocks as the bottleneck. 2D convolution filters are used for feature extraction since the input is similar to an image. It is worth mentioning that we use upsampling function followed by the same convolution instead of a deconvolution filter to avoid checkerboard artifacts (Odena et al., 2016). The discriminator is a PatchGAN, namely a fully-convolutional structure. Equalized learning rate layer and pixel-wise norm layer are applied to improve the stability of training (Karras et al., 2018). Leaky ReLU (Maas et al., 2013) is applied as the activation function.

2.3.4. Correlative spectral attention

To capture the latent correlation between the input EEG segment and the target SEEG segment, the CSA is proposed. It adopts a mutual attention module to model the correlation from a perspective of magnitude spectra. The mutual attention module serves as an observer that learns to express to what extent the correlation between EEG and SEEG segments is preserved and then passes the message to the discriminator as a supply.

Specifically, given an input pair (e, s), only the magnitude E_{mag} and $S_{mag} \in \mathbb{R}^{m\times n}$ are extracted. The magnitude is defined as a sequence of frequency vectors $(f_1, f_2, ..., f_m)$ where each frequency vector $f_i \in \mathbb{R}^n$ is a time series $(t_1, t_2, ..., t_n)$. To map

TABLE 1 Architectures of the proposed generator and discriminator.

Layers	Output shape	Norm./Act.
Generator		
Input	2 × 128 × 128	–
Padding	2 × 128 × 128	ELR/LReLU/PN
Conv2d	32 × 64 × 64	ELR/LReLU/PN
Conv2d	64 × 32 × 32	ELR/LReLU/PN
ResBlock*2	64 × 32 × 32	ELR/LReLU/PN
Upsample	64 × 64 × 64	–
Conv2d	32 × 64 × 64	ELR/LReLU/PN
Upsample	32 × 128 × 128	–
Conv2d	16 × 128 × 128	ELR/LReLU/PN
Padding	16 × 134 × 134	–
Conv2d	2 × 128 × 128	Tanh
Discriminator		
Input	4 × 128 × 128	–
Conv2d	16 × 64 × 64	ELR/LReLU
Conv2d	32 × 32 × 32	ELR/LReLU
Conv2d	64 × 16 × 16	ELR/LReLU
Conv2d	128 × 15 × 15	ELR/LReLU
Conv2d	1 × 14 × 14	–

E_{mag} and S_{mag} to a proper space, learned affine transformations A_1 and A_2 are applied. $A(\cdot, w, b)$ is defined as

$$A(X, w, b) = X \cdot w + b \tag{4}$$

where X is the input matrix, $w \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n$ are learnable affine parameters. To further obtain the correlative expressions, mutual attention is calculated between E_{mag} and S_{mag}. The attention score α is defined as

$$\alpha = softmax(\frac{A_1(E_{mag}, w_1, b_1) \cdot A_2(S_{mag}, w_2, b_2)^{\mathrm{T}}}{\sqrt{n}}) \tag{5}$$

where each scalar in $\alpha \in \mathbb{R}^m$ can be considered as the level of correlations between different combinations of EEG and SEEG frequencies. E_{mag} is multiplied by α to obtain the output of CSA

$$CSA(e, s) = \alpha \cdot E_{mag} \tag{6}$$

and concatenate it with the original pair (e, s) as the final input to D. Here, the output of CSA acts as an auxiliary supervision signal which suggests the principal components of input correlated to the target. CSA reinforces the temporal-frequency representation to be better utilized by the downstream PatchGAN.

2.3.5. Weighted patch prediction

The patch-based technique described in Isola et al. (2017) is not powerful enough to discriminate the features of an image in low variance situations since a little change in features provides many similar sub-images of monotonous mode and therefore leads to mode collapse. In practice, the low variance in IF along each frequency is observed, which leads to mode collapse in both generated magnitude and IF spectra. The collapsed spectra have a salient characteristic of repetitive stripes along with each frequency. To eliminate the undesirable results, a patch-weighting technique is adopted by applying a learnable weight to the prediction of a patch. The weighted patch predictions are then conforming to a distorted distribution distinguishable from each other. In addition, the averaged representation of all patch predictions is weighted to provide a global view of the input to adjust the final prediction.

Specifically, the original PatchGAN D outputs a one-channel matrix $P \in \mathbb{R}^{1\times k\times k}$, of which the element represents the prediction of a patch. k is the number of rows in the output matrix. P is flattened to obtain vector $p \in \mathbb{R}^{k^2}$. The global representation p_{global} is calculated by averaging the predictions in p and applying a non-linear transformation:

$$p_{global} = LReLU(\frac{1}{k^2}\sum_{i=1}^{k^2} p_i) \tag{7}$$

We choose leaky rectified activation (LReLU) (Maas et al., 2013) as the non-linear function because it has been proved to work well for training GAN models (Radford et al., 2015). LReLU helps to ensure the gradient can flow through the entire architecture. Then, p_{global} and p are concatenated in preparation for the weighted prediction $\widehat{y}$, which is defined as

$$q = cat(p_{global}, p), q \in \mathbb{R}^{k^2+1} \tag{8}$$

$$\widehat{y} = \sum_{i=1}^{k^2+1} \beta_i q_i \tag{9}$$

In Equation (9), β_i denotes the learnable weight assigned to the global view and the patch predictions. The behavior of weighting disturbs the original monotonous distribution among patches. The global representation p_{global} plays a role in providing the complete insight into the input since it collects views from all patches.

2.3.6. Objectives

WGAN-GP (Gulrajani et al., 2017) is adopted to stabilize the training of cGAN. The objective with the proposed auxiliary modules are formulated as follows:

$$\begin{aligned}\mathcal{L}_{GAN} = & \mathbb{E}_{e,s\sim p_r(e,s)}[D(e, s, CSA(e, s))] - \\ & \mathbb{E}_{e\sim p_r(e)}[D(e, G(e), CSA(e, G(e)))] + \\ & \lambda_{gp}\mathbb{E}_{\hat{e},\hat{s}\sim p_i(\hat{e},\hat{s})}[(\|\nabla_{(\hat{e},\hat{s})}D(\hat{e}, \hat{s}, CSA(\hat{e}, \hat{s}))\|_2 - 1)^2]\end{aligned} \tag{10}$$

where $p_r(e, s)$ is the joint distribution of real pairs and $p_r(e)$ is the distribution of EEG. $p_i(\hat{e}, \hat{s})$ is the joint distribution defined on the interpolated space derived from real and synthesized pairs. WGAN-GP uses the interpolated space derived from the real and synthesized samples to represent the input space of D in order to compute gradients. λ_{gp} is used to scale the gradient penalty.

L1 loss is adopted in the generator for it contributes to capturing the low-frequency components of a image (Isola et al., 2017), which in this work corresponds to the slow-changing regions of spectra:

$$\mathcal{L}_{L1} = \mathbb{E}_{e,s \sim p_r(e,s)}[\|s - G(e)\|_1] \tag{11}$$

As a result, the final min-max optimization objective is

$$\mathcal{G}^*_{E2SGAN} = \arg\min_G \max_D \mathcal{L}_{GAN} + \lambda_{L1}\mathcal{L}_{L1} \tag{12}$$

where λ_{L1} is a hyper-parameter adjusting the L1 loss.

3. Results

3.1. Subjects and data recording

Seven subjects participated in the study. Subjects were patients with intractable epilepsy who had implanted SEEG electrodes for pre-surgical assessment of their seizure focus. All implant parameters were determined solely by clinical needs rather than those of this research. SEEG signals were acquired using a clinical recording system (EEG-1200C, Nihon Kohden, Irvine, CA) and sampled at 2,000 Hz. We also recorded scalp EEG simultaneously. All subjects gave informed consent to this study, which was approved by the Ethics Committee of Huashan Hospital (Shanghai, China).

3.2. Preprocessing

Both EEG and SEEG signals are vulnerable to noise interference. Slow signal drifts mask genuine cortical activity in the low frequency range (de Cheveigné and Arzounian, 2018). Electrical line noise causes unwanted effects on a fixed frequency of 50 or 60 Hz and their harmonics. To remove these unwanted components from the signal, EEG signal at 1 Hz and SEEG at 0.5 Hz is high-pass filtered to remove slow signal drifts (Li et al., 2018), and the signal at 50 Hz and their harmonics are notch filtered to remove line noise. SEEG electrodes located in the cortex are filtered out for the reason that they contributed little to the scalp signal. For the convenience of computation, the filtered signal is down-sampled to 64 Hz where the majority of power is distributed. Segmentation was performed on the continuous brain signals. A sliding window with a size of 1,016 sampling points (15.875 s) and a stride of a quarter of the window size is performed on the filtered signal. STFT with the window size of 256 and the hop length of 8 was performed to transform each segment to magnitude and phase spectra matrix with the same shape of (128, 128). IF spectra matrix from phase spectra is then derived. The final representation of each segment was a (2, 128, and 128) tensor.

We used pre-ictal EEG and SEEG recorded simultaneously from seven epileptic rest-state patients. The placement of EEG electrodes conformed to the 10–20 system. After segmentation and pairing, each subject had 8,000 EEG-SEEG pairs as the training set, 200 as the validation set, and 200 as the test set.

3.3. Evaluation metrics

The quality of results is evaluated by calculating the distance between real and generated SEEG segments from both temporal and frequency perspectives. From a temporal perspective, dynamic time warping (DTW) is used which matches two-time series through a dynamic programming strategy. From a frequency perspective, the Hellinger Distance and root mean square errors (RMSE) of the power spectral density (PSD) are adopted. The Hellinger Distance reflects the similarity of the power distribution of different frequencies. MSE of PSD reflects to what degree the dominant components of power are recovered. Inception score (Salimans et al., 2016) and Frechet Inception Distance (Heusel et al., 2017) are no longer considered in this work which is a determined mapping instead of generation from noise.

For evaluation, the leave-one-patient-out cross-validation strategy is employed. The averaged results are computed only on the test sets. Evaluation of the input EEG was used as a baseline. The evaluations are post-processed according to Equation (13) by dividing them by the baseline evaluation. A binary logarithm was taken as the final result. For each experiment, the baseline evaluation is shown on the header of the tables. For all metrics, a smaller value indicates a better result.

$$Eval_{log}(S_{fake}, S_{real}) = \log_2 \frac{Eval(S_{fake}, S_{real})}{Eval(E_{real}, S_{real})} \tag{13}$$

3.4. Baselines

The proposed architecture is compared with different approaches using mainstream generative models. ASAE and AE adopt an encoder-decoder architecture while EEGGAN, GANSynth, and Pix2pix are based on GANs.

3.4.1. ASAE

Asymmetric-Symmetric Autoencoder uses an asymmetric autoencoder to map scalp signal to intracranial signal, followed by a symmetric autoencoder to enhance the generated results.

3.4.2. AE

This baseline has the same architecture as the generator in our method (see Table 1).

3.4.3. EEGGAN

EEGGAN uses an improved WGAN-GP by gradually relaxing the gradient constraint to map a noise distribution to the EEG distribution. In the experiments, the input noise is replaced with EEG segments to generate SEEG.

3.4.4. uTSGAN

uTSGAN uses two WGAN-GP sub-models to interactively learn the time- and frequency-domain generation of time series from a noise distribution. The input noise is replaced with reshaped EEG segments. It has to be mentioned that since the authors of uTSGAN did not open-source their code, we only reproduce their work in a simple way as efficiently as possible. We use 2D Conv for the spectral WGAN and 1D Conv for the temporal WGAN. The loss function and optimizers follow as described in the original work (Smith and Smith, 2020, 2021).

3.4.5. Pix2pix

Pix2pix is a generic image-to-image translation approach based on cGANs. In the experiments, we implement the generator with the same architecture as the AE method and adopt PatchGAN which was originally used in the work (Isola et al., 2017).

3.4.6. GANSynth

GANSynth is a GAN-based method used in audio synthesis which takes the magnitude and IF spectra as input to generate the audio segments in a progressive way. The input noise is also replaced with EEG segments.

For the baselines that are originally applied to signal generation such as ASAE, EEGGAN, and GANSynth, we follow the setting specified in the paper. In other words, STFT will not be performed to preserve the temporal representation for ASAE and EEGAN while magnitude and IF spectra are used for GANSynth. For the rest, magnitude and phase (instead of IF) spectra are used as a two-channel image to be compatible with the CNN-based architectures. For uTSGAN, we use both the temporal representation and the magnitude matrix.

3.5. Experiment settings

λ_{gp} and λ_{L1} are set to 10. The learning rate of G and D is set to 1e-5 and 2e-5, respectively. In the training phase, the model is first trained by 80 epochs without the CSA module and then jointly trained by another 40 epochs. We adopted Adam optimizer to perform gradient descent optimization (Kingma and Ba, 2014) and implemented our framework with Pytorch. Preprocessing is done with the help of the MNE toolkit (Gramfort et al., 2013).

3.6. Performance of different methods

3.6.1. Comparison with baseline methods

The results of the proposed method trained on the aligned dataset guided by the proposed matching strategy introduced in Section 2.2 are denoted as E2SGAN-S. The comparison results are shown in Table 2. The full visual results are provided in Supplementary Figures 1, 2. ASAE performs poorly in DTW and HD because its encoder-decoder architecture simply based on stacks of fully-connected layers is not powerful enough to learn the complicated mapping between EEG and SEEG. Although it seems to perform well on PSD, the fact is that the model only has learned low-frequency pulses that dominate the magnitude of the power spectrum. AE is able to capture the majority of low-frequency features but underfits the high frequency. This reflects on the temporal result as slow fluctuations with simple ripples. Both the autoencoder-based methods overfit the low frequencies because of no extra regularization on high-frequency features. EEGGAN learns the mapping directly from the temporal domain and captures the general distribution of the target power spectrum, but it fails to learn the exact features of the waveform, which leads to a competitive performance only on HD. uTSGAN is inferior to the proposed E2SGAN and other baselines. Pix2pix has a slightly better performance compared to AE on PSD and HD. The patch-based discriminator enables it to learn more complicated patterns but also makes it hard to generate them accurately, resulting in poor performance on DTW. GANSynth achieves the most competitive performance among baselines because it utilizes IF spectra which contributes to the accuracy in phase recovery. The progressive approach enables it to capture the general power distribution and thus perform well on HD metrics. The proposed method achieves the best results on all metrics. The DTW result far superior to the others demonstrates the effectiveness of adopting IF spectra and the WPP technique. In addition, the lowest MSE of PSD and HD results suggest that the CSA module successfully captures not only the general power distribution of the target but also the exact magnitude of different frequencies by imposing regularization on the correlations between input and target frequencies.

3.6.2. Influence of matching strategies

Furthermore, the performance of the proposed model on datasets constructed under the guidance of different matching

TABLE 2 Performance comparison with baselines.

Method	DTW/1.636e-3	PSD/1.128e-9	HD/2.886e-6
ASAE (Antoniades et al., 2018)	2.867 ± 0.126	−1.384 ± 1.794	0.490 ± 0.800
AE$^{(*)}$	−0.059 ± 0.815	−1.357 ± 1.659	0.337 ± 0.756
EEGGAN (Hartmann et al., 2018)	−0.019 ± 0.624	−0.821 ± 1.457	0.068 ± 0.734
uTSGAN (Smith and Smith, 2021)	3.867 ± 0.012	3.424 ± 0.134	3.127 ± 0.050
Pix2pix (Isola et al., 2017)	0.133 ± 0.829	−1.396 ± 1.754	0.316 ± 0.834
GANSynth (Engel et al., 2019)	−0.107 ± 0.740	−0.929 ± 1.453	0.191 ± 0.763
E2SGAN-N	**−0.424 ± 0.400**	0.185 ± 0.298	0.572 ± 0.565
E2SGAN-S	−0.414 ± 0.764	**−1.480 ± 1.609**	**−0.221 ± 0.843**

AE$^{(*)}$ is implemented with the same architecture as the generator of the proposed method. The bold values indicate the baseline values of evaluation.

TABLE 3 Ablation study.

Method	DTW/1.636e-3	PSD/1.128e-9	HD/2.886e-6
w/o CSA & WPP	−0.318 ± 0.723	−1.633 ± 1.732	−0.393 ± 0.948
w/o WPP	−0.325 ± 0.729	**−1.681 ± 1.734**	**−0.425 ± 0.953**
w/o CSA	−0.346 ± 0.770	−1.385 ± 1.615	−0.112 ± 0.851
E2SGAN	**−0.414 ± 0.764**	−1.480 ± 1.609	−0.221 ± 0.843

The bold values indicate the baseline values of evaluation.

strategies is investigated. A variant E2SGAN-N is proposed to be trained on the dataset guided by the nearest neighbor matching strategy. We adopt a most intuitive and practical strategy, that is to match the EEG and SEEG channels with the closest physical distance, and ensure the selection is not repeated. The results are shown at the bottom of Table 2. E2SGAN-N and E2SGAN-S have similar performance in the time domain, but E2SGAN-N is inferior to most baselines in the frequency domain. The main reason is that our proposed matching strategy explicitly guarantees the correlation between EEG-SEEG. This finding confirms the point mentioned in the previous Section 2.2, that the signals recorded by channels with closer distance do not necessarily have obvious correlations in the frequency domain. This can be mainly attributed to the irregular topology of gyri and sulci in the brain so that the propagation direction of intracranial signals does not necessarily follow the direction closest to the scalp. Assuming that the proposed model is sensitive to EEG-SEEG correlation, a paired dataset with a stronger correlation should achieve better performance. This experiment demonstrates the ability of E2SGAN to implicitly capture the correlation between EEG and SEEG.

3.7. Ablation study

In order to verify how CSA and WPP are beneficial to the whole model, three variants are compared to E2SGAN. For the "w/o WPP" variant, the WPP module is removed and the original PatchGAN prediction is used as output. For "w/o CSA" variant, we do not concatenate the CSA output to the input of PatchGAN. For "w/o CSA & WPP" variant, both of the aforementioned changes are adopted. As shown in Table 3, the removal of "WPP" degrades the performance in the temporal domain (DTW) and the removal of "CSA" degrades the performance in the frequency domain (PSD & HD). Although the two variants without "WPP" seem to perform better in the frequency domain, they actually suffer from mode collapse, leading to poor temporal robustness.

3.7.1. Temporal robustness study

We further compare the proposed method to w/o CSA and w/o CSA & WPP in the temporal domain to verify the temporal robustness by calculating the standard deviation (STD) distribution of the RMSE between the real and synthesized SEEG. Specifically, each SEEG segment is equally divided into sub-segments, the width of which is traversed from 2 to half of the whole length. RMSE is calculated gradually on sub-segments, and the STD distribution of it on different scales of width is plotted in Figure 5. The horizontal axis is STD and the vertical axis is density. It can be clearly observed that the majority of STD of both methods with WPP are centrally distributed in low-value regions. The high STD in w/o CSA and WPP method implies some kind of spur occurring in the generated segments, which is actually observed in visual results.

3.8. Input-signal perturbation

To determine what key features in the input signal play a significant role in relating EEG to the simultaneous SEEG, input-signal perturbation is performed as is described in Gemein et al. (2020) and Schirrmeister et al. (2017). Specifically, random noise $z \sim N(0, \sigma^2)$ drawn from Gaussian distribution (with mean 0 and standard deviation σ identical with that of the original input) was added to the magnitude spectra of input EEG e while

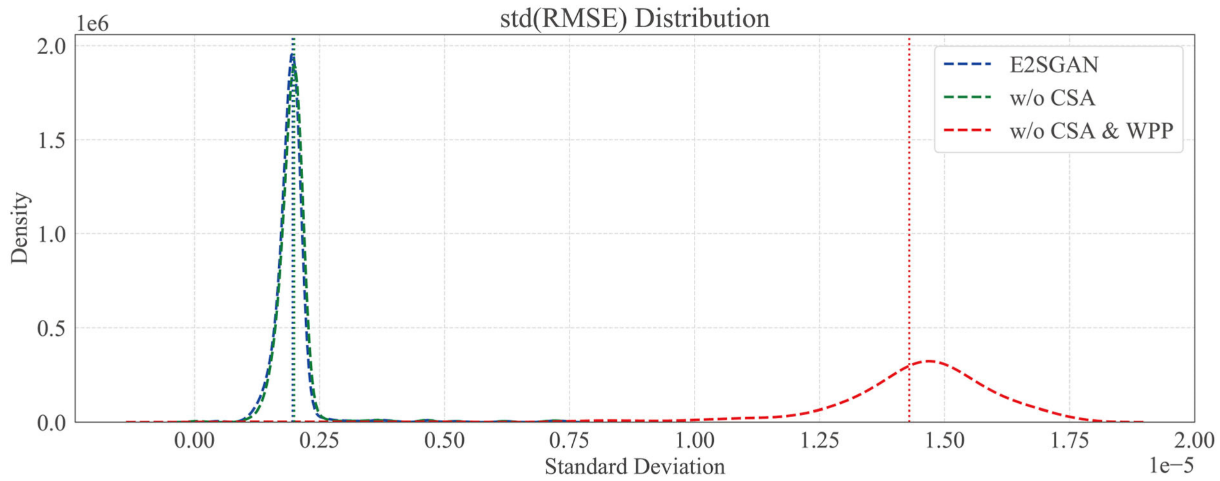


FIGURE 5
Temporal robustness study by comparing the standard deviation (SD) distribution of RMSE. Dotted vertical lines are the means of SD.

the phase was kept unperturbed. RMSE of PSD is computed with the perturbed EEG $\hat{e}$ as input. Changes compared to the unperturbed result were computed to indicate the effect Δ_{effect} of perturbation.

$$\Delta_{effect} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(PSD(G(\hat{e}))^{(i)} - PSD(s)^{(i)})^2} - \sqrt{\frac{1}{m}\sum_{j=1}^{m}(PSD(G(e))^{(j)} - PSD(s)^{(j)})^2} \tag{14}$$

Δ_{mag} denoting the changes in the magnitude of the input EEG was obtained by averaging the added noise z. Such perturbation was done on every EEG-SEEG pair and repeated 50 times on each investigated feature, including four frequency bands: δ (0–4 Hz), θ (4–7 Hz), α (8–15 Hz), β (16–31 Hz). For each feature, Δ_{effect} is correlated with Δ_{mag} by computing the correlation coefficient. It could be determined whether the increase or decrease in the magnitude of the investigated features contributed to a better or worse SEEG synthesis.

The perturbation operation is carried out individually on each patient. We selected three patients as representatives, who showed three kinds of responses to the proposed model to perturbation: (1) the captured perturbation-sensitive area is consistent with the epileptic focus; (2) the captured sensitive area is close to the focus; (3) being unable to capture sensitive areas near the focus due to data missing. The three cases are displayed in Figure 6 by visualizing the topological map of the correlation coefficients. EEG electrodes are partly missing due to the restriction in the collection and the proposed pairing strategy (which causes the unmatched channels to be discarded). We pay attention to the sensitivity of signals at different electrode locations to perturbation in different frequency bands. In patient 1, a dramatic effect is found in the parietal and temporal zones (*P3* and *T5*) of all bands and is most significant in the δ band. In patient 2, the most affected areas are the left frontal and central zones (*F3*, *C3*). In patient 3, the left temporal zone (*T3*, *T5*) is more sensitive. The added perturbation can be seen as a simulation of the underlying micro-abnormal discharges that occur before seizures. To verify that the sensitive channels we find are clinically significant, the seizure location provided by the clinician is marked with yellow-edged dots. Patient 1 shows that the seizure location (*P3*) coincides with the sensitive zone. For patient 2, the seizure location (*T3*, *F7*) is close to the sensitive zone we have found despite the missing of channels. An exceptional case is patient 3, where the real epileptic focus (*T4*) is in the right temporal zone, far away from the available channels in the training dataset. The cases of patients 1 and 2 demonstrate that the proposed model is able to capture the abnormal signal before seizures and is preserved in the synthesized SEEG. If the seizure location is too far away from the electrodes available in the dataset, our model will perform poorly. We provide the rest of the patients with Supplementary materials.

It has been verified that the δ power is associated with sleep and shows abnormal changes prior to seizures (Minecan et al., 2002; Long et al., 2021). The perturbation results in the δ band confirm the expected conclusion.

3.9. Case study of visual results

A visual case study is conducted by comparing the spectral and temporal results synthesized by the proposed method to the groundtruth. As is shown in the top rows of Figure 7, the proposed method produces clear results in both frequency and IF spectra, which indicates it has captured the details of how

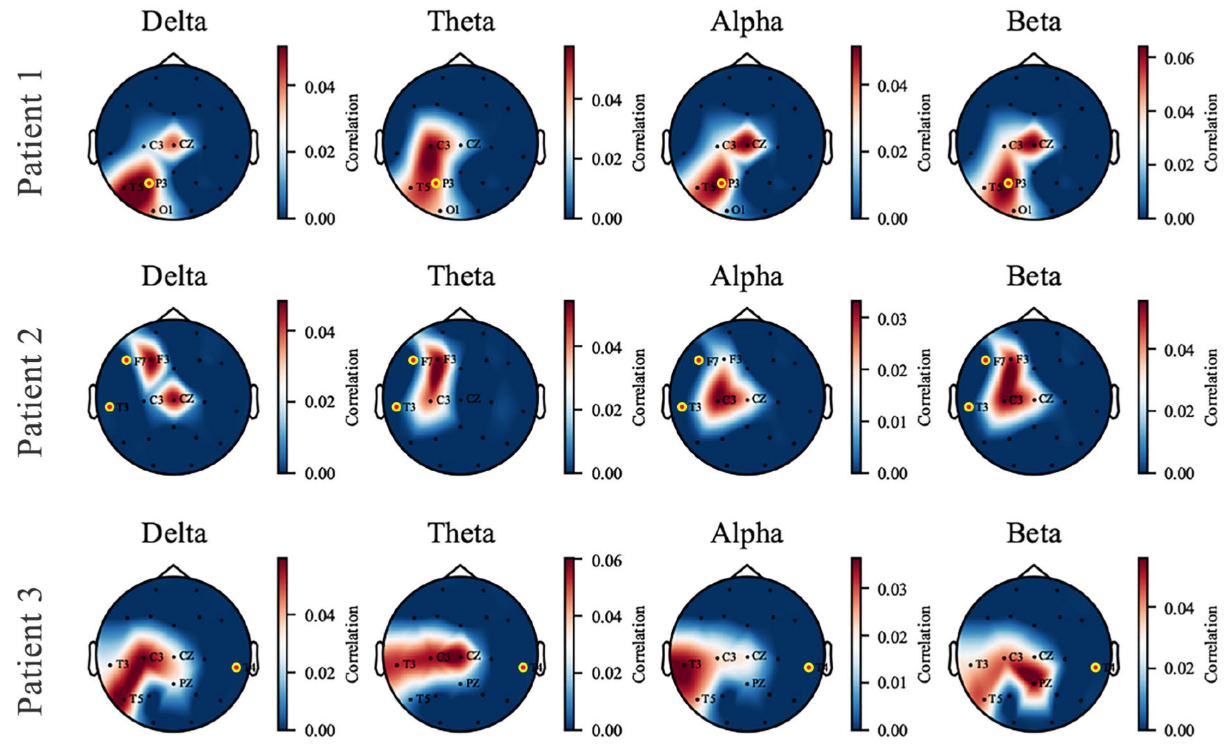


FIGURE 6
The topological map of correlation coefficient from input-signal perturbation where red indicates a high correlation and blue indicates a low correlation. Four bands (δ, θ, α, and β) are investigated on each patient, and three of them are displayed as representatives. The names of available EEG channels for each patient are listed, and the channels nearby a seizure focus are marked with yellow-edged dots. Patients 1 and 2 are good examples that the sensitive zone (left parietal and temporal zone for Patient 1, left frontal and temporal for Patient 2) coincides with the seizure location, while Patient 3 is an exceptional case due to the missing of *T4* channel in the dataset.

the power and phase are distributed along with time. From the bottom rows, it can be seen that the proposed method is able to produce the general morphology highly close to the groundtruth although the details of ripples are still far from satisfactory.

4. Discussion

In this paper, a GAN-based framework is introduced for the task of EEG-to-SEEG translation. First, a matching strategy is developed to select EEG-SEEG pairs. Second, E2SGAN is proposed to learn from the magnitude and IF spectra to synthesize the simultaneous SEEG segment given an input EEG segment. CSA and WPP technique are proposed to give a further boost to the discriminator. Extensive comparison experiments have demonstrated the capability of the proposed framework to transform EEG segments into SEEG segments. To find out whether the model has captured clinically significant features, a perturbation experiment is conducted. The final result shows that the synthesized SEEG signal retains the abnormal discharges before seizures.

4.1. Why not directly model the physical distance between EEG and SEEG?

Intuitively, it is more reasonable to generate the nearest EEG-SEEG pairs. We favor the impact of physical distance on SEEG generation, but it is not the only factor. In this work, we have tried to match EEG-SEEG pairs using a strategy based on the nearest physical distance. However, we found that EEG-SEEG pairs based solely on the physical distance were not necessarily strongly correlated (see Section 2.2 and Figure 3). This observation has also been confirmed in relevant clinical studies, where the geometry of the brain and its complex interconnections affect signal propagation (Frauscher et al., 2018). In the comparison experiment, we also experimented with the nearest physical distance matching strategy, but the results were not satisfactory. Therefore, we incorporated the HD distance into the matching strategy to limit the correlation between an EEG-SEEG pair. In this way, we avoid the influence of some unrelated pairs in the training process, even if they have a relatively close physical distance.

We agree that modeling the physical distance into the model is a worthwhile and practical consideration, and there have been

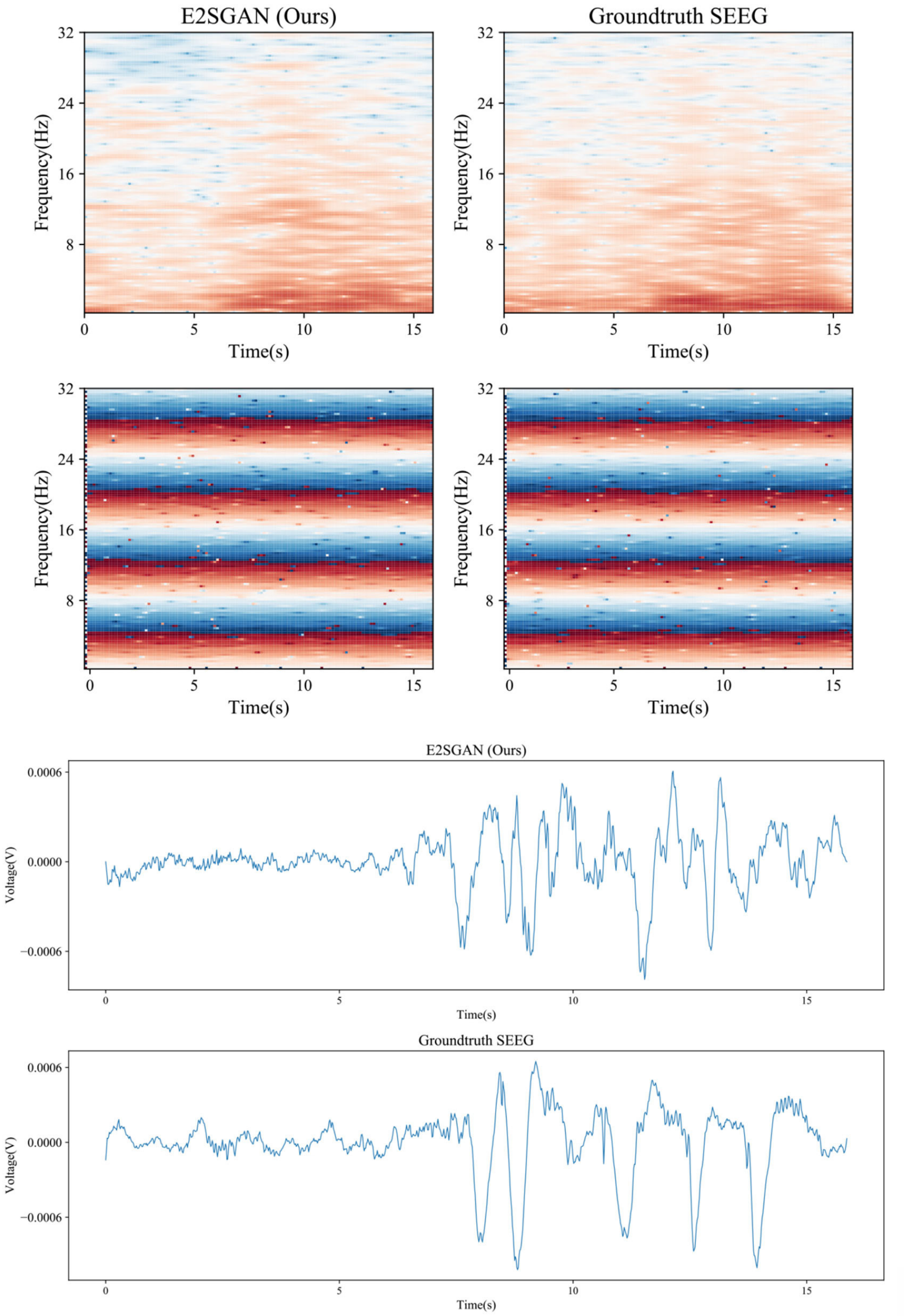


FIGURE 7
Case study of spectral results **(top)** and temporal results **(bottom)** synthesized by the proposed method compared to the groundtruth.

many related works to spatially model the fixed position of EEG (Fang et al., 2020; Jia et al., 2021). However, for SEEG electrodes, this can be a huge challenge. SEEG positions are not fixed and cannot be uniformly modeled for all patients. Additionally, due to the very limited number of patients with simultaneous SEEG data we have access to, it is difficult to learn location representation with such large variance across different patients from a data-driven perspective. Another approach is to learn a fixed position representation and train a model per patient. Although this is feasible, the next problem we cannot avoid is that the position space of SEEG is continuous. For the generation task, we cannot achieve the goal of generating the accurate SEEG segment given an arbitrary position in the continuous space. To this end, the proposed matching strategy can essentially be viewed as discretizing the continuous position space (by dividing the physical distance into intervals) and selecting the "nearest" electrode pair that guarantees EEG-SEEG correlation. In other words, we incorporate physical distance as a prior knowledge into the data preprocessing process. In future work, we aim to collect a larger dataset and try to model the electrode position, distance, direction, and other information in a unified way through the deep-learning network to explore whether the existing technology can fully capture the rich clinical information.

4.2. Trade-off between extracted features and raw signal

When analyzing EEG signals, clinicians can judge whether there are signs of seizures through either the raw waveform and spectrum or the meaningful features computed by mathematical methods. Different input types affect the difficulty of generation tasks.

If the extracted features are used for the generation, the generated results can be directly applied to specific downstream tasks (Luo Y. et al., 2020). Traditional signal features such as Differential Entropy (DE) and PSD are highly informative and discriminative for specific tasks such as emotion recognition. They can be easily learned by neural networks and generate accurate results that are suitable for downstream tasks. However, the high compatibility with downstream tasks also limits the applicability of the generated results. For each feature or task, a neural network needs to be trained separately, which is not feasible in clinical use.

If the raw signal is directly used as the target, the generated results are applicable to a wider range of scenarios. Ideally, the generated results would approximate the real signal distribution, on which clinicians can perform any posterior analysis of interest. However, this type of generation is tricky because the informative signal components are often sparse and difficult for the network to fit.

There is a trade-off between using the extracted features or the raw signal as the generation target. Hence, large-scale simultaneously recorded EEG-SEEG datasets are necessary, which can improve the performance of end-to-end generation tasks to a certain extent. In addition, appending a downstream task to the generation stage that recognizes specific features can further improve the generation results, while ensuring good adaptability to specific tasks at the same time.

4.3. Challenges of different channel mapping assumptions

So far, only the one-to-one channel mapping has been considered in this work. Guided by this assumption, the proposed method can theoretically capture arbitrary EEG-SEEG correlations as long as the available dataset covers adequate EEG-SEEG combinations and achieve the brain-wide generative capability. In practice, this mapping assumption is not always plausible due to complex electrode arrangements. There are situations where it is impossible to find electrodes that fit the distance required by the proposed strategy, such as those positioned nearby wounds. Moreover, the scarcity of patients with synchronized EEG recordings is still a hindrance.

A more realistic assumption than one-to-one modeling is multiple-to-one/multiple-to-multiple mapping. EEG, as the collection of SEEG on the scalp, is actually the superposition of multiple intracranial discharge sources. However, multiple-to-one/multiple-to-multiple mapping is a more intractable problem that requires elaborate modeling techniques to extract the universal relations among multiple heterogeneous channels. For clinical research based on EEG signals, the study of multiple-to-multiple mapping is of great significance because the interpretive work based on it can reveal how intracranial signals propagate, thereby assisting clinicians to solve the problem of source localization. In the future, graph modeling using GCN may be more suitable for tasks based on this assumption, and graph-based interpretation algorithms will reveal richer principles of intracranial EEG signal propagation.

For future work, our proposed method can serve as a baseline for solutions developed on massive datasets or as a benchmark for formulating the multiple-to-one/multiple-to-multiple assumptions.

4.4. Unsupervised learning in SEEG generation

In this research, the adopted GANs framework is an unsupervised learning paradigm. Unsupervised learning eliminates the stage of labeling sample categories by experts and gets rid of the constraints of limited data sets. This property is particularly valuable because clinical data is often massive and unlabeled, and qualified clinicians rarely have the chance to withdraw from clinical work. With the advent of unsupervised learning, a new paradigm of pre-training has also emerged and has been applied in EEG research (Kostas et al., 2021; Yue et al., 2022; Zhang et al., 2022). This new paradigm inspired us to see if pre-trained models trained on large-scale EEG datasets can be used to generate SEEG signals after a fine-tuning stage. This is also one of the ideal solutions for the research topic in this paper.

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 Institutional Review Board, Huashan Hospital, Fudan University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

MH, JC, WJ, and XW designed the proposed framework and drafted the manuscript. SJ, SM, and LC collected the data and analyzed the experiment results. All authors contributed critically to this work and approved of the upcoming edition.

Funding

This work was supported by NSFC grants (No. 62136002 and 61972155) and the Science and Technology Commission of Shanghai Municipality (20DZ1100300).

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.971829/full#supplementary-material

References

Antoniades, A., Spyrou, L., Martin-Lopez, D., Valentin, A., Alarcon, G., Sanei, S., et al. (2018). Deep neural architectures for mapping scalp to intracranial EEG. *Int. J. Neural Syst.* 28:1850009. doi: 10.1142/S0129065718500090

Cao, M., Galvis, D., Vogrin, S. J., Woods, W. P., Vogrin, S. J., Wang, F., et al. (2022). Virtual intracranial EEG signals reconstructed from meg with potential for epilepsy surgery. *Nat. Commun.* 13:994. doi: 10.1038/s41467-022-28640-x

Chabardes, S., Abel, T. J., Cardinale, F., and Kahane, P. (2018). Commentary: understanding stereoelectroencephalography: what's next? *Neurosurgery* 82, E15–E16. doi: 10.1093/neuros/nyx499

Chen, G., Lu, G., Xie, Z., and Shang, W. (2020). Anomaly detection in EEG signals: a case study on similarity measure. *Comput. Intell. Neurosci.* 2020:6925107. doi: 10.1155/2020/6925107

Cossu, M., Cardinale, F., Castana, L., Citterio, A., Francione, S., Tassi, L., Benabid, A. L., and Lo Russo, G. (2005). Stereoelectroencephalography in the presurgical evaluation of focal epilepsy: a retrospective analysis of 215 procedures. *Neurosurgery* 57, 706–718. doi: 10.1227/01.NEU.0000176656.33523.1e

de Cheveigné, A., and Arzounian, D. (2018). Robust detrending, rereferencing, outlier detection, and inpainting for multichannel data. *NeuroImage* 172, 903–912. doi: 10.1016/j.neuroimage.2018.01.035

Engel, J. H., Agrawal, K. K., Chen, S., Gulrajani, I., Donahue, C., and Roberts, A. (2019). "Gansynth: adversarial neural audio synthesis," in *7th International Conference on Learning Representations, ICLR 2019* (New Orleans, LA).

Fang, Z., Wang, W., Ren, S., Wang, J., Shi, W., Liang, X., et al. (2020). "Learning regional attention convolutional neural network for motion intention recognition based on EEG data," in *IJCAI.* (Yokohama), 1570–1576. doi: 10.24963/ijcai.2020/218

Frauscher, B., von Ellenrieder, N., Zelmann, R., Dolezalova, I., Minotti, L., Olivier, A., et al. (2018). Atlas of the normal intracranial electroencephalogram: neurophysiological awake activity in different cortical areas. *Brain* 141, 1130–1144. doi: 10.1093/brain/awy035

Gemein, L. A. W., Schirrmeister, R. T., Chrabaszcz, P., Wilson, D., Boedecker, J., Schulze-Bonhage, A., et al. (2020). Machine-learning-based diagnostics of EEG pathology. *NeuroImage* 220:117021. doi: 10.1016/j.neuroimage.2020.117021

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014*, eds Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (Montreal, QC), 2672–2680.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2013). MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* 7:267. doi: 10.3389/fnins.2013.00267

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. (2017). Improved training of wasserstein gans. *arXiv preprint arXiv:1704.00028.* doi: 10.48550/arXiv.1704.00028

Hartmann, K. G., Schirrmeister, R. T., and Ball, T. (2018). Eeg-gan: Generative adversarial networks for electroencephalograhic (EEG) brain signals. *arXiv preprint arXiv:1806.01875.* doi: 10.48550/arXiv.1806.01875

He, K., Wan, S., Sheng, J., Liu, D., Wang, C., Li, D., et al. (2019). A high-performance compact magnetic shield for optically pumped magnetometer-based magnetoencephalography. *Rev. Sci. Instrum.* 90:064102. doi: 10.1063/1.5066250

Henry, J. C. (2006). Electroencephalography: basic principles, clinical applications, and related fields. *Neurology* 67, 2092–2092. doi: 10.1212/01.wnl.0000243257.85592.9a

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. (2017). "GANs trained by a two time-scale update rule converge to a local NASH equilibrium," in *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017*, eds I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, and R. Garnett (Long Beach, CA), 6626–6637.

Isola, P., Zhu, J., Zhou, T., and Efros, A. A. (2017). "Image-to-image translation with conditional adversarial networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* (Honolulu, HI: IEEE Computer Society), 5967–5976. doi: 10.1109/CVPR.2017.632

Jia, Z., Lin, Y., Wang, J., Ning, X., He, Y., Zhou, R., et al. (2021). Multi-view spatial-temporal graph convolutional networks with domain generalization for sleep stage classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* 29, 1977–1986. doi: 10.1109/TNSRE.2021.3110665

Jiao, Z., You, H., Yang, F., Li, X., Zhang, H., and Shen, D. (2019). "Decoding EEG by visual-guided deep neural networks," in *IJCAI* (Macao), 1387–1393. doi: 10.24963/ijcai.2019/192

Karras, T., Aila, T., Laine, S., and Lehtinen, J. (2018). "Progressive growing of GANs for improved quality, stability, and variation," in *6th International Conference on Learning Representations, ICLR 2018* (Vancouver, BC).

Kaur, K., Shih, J. J., and Krusienski, D. J. (2014). Empirical models of scalp-EEG responses using non-concurrent intracranial responses. *J. Neural Eng.* 11:035012. doi: 10.1088/1741-2560/11/3/035012

Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980.* doi: 10.48550/arXiv.1412.6980

Kostas, D., Aroca-Ouellette, S., and Rudzicz, F. (2021). BENDR: using transformers and a contrastive self-supervised learning task to learn from massive amounts of EEG data. *CoRR, abs/2101.12037.* doi: 10.3389/fnhum.2021.653659

Li, G., Jiang, S., Paraskevopoulou, S. E., Wang, M., Xu, Y., Wu, Z., et al. (2018). Optimal referencing for stereo-electroencephalographic (SEEG) recordings. *NeuroImage* 183, 327–335. doi: 10.1016/j.neuroimage.2018.08.020

Li, R., Wang, Y., and Lu, B. (2021). "A multi-domain adaptive graph convolutional network for EEG-based emotion recognition," in *MM '21: ACM Multimedia Conference, Virtual Event*, eds H. T. Shen, Y. Zhuang, J. R. Smith, Y. Yang, P. Cesar, F. Metze, and B. Prabhakaran (Chengdu), 5565–5573.

Long, S., Ding, R., Wang, J., Yu, Y., Lu, J., and Yao, D. (2021). Sleep quality and electroencephalogram delta power. *Front. Neurosci.* 15:803507. doi: 10.3389/fnins.2021.803507

Luo, T.-j., Fan, Y., Chen, L., Guo, G., and Zhou, C. (2020). Eeg signal reconstruction using a generative adversarial network with Wasserstein distance and temporal-spatial-frequency loss. *Front. Neuroinform.* 14:15. doi: 10.3389/fninf.2020.00015

Luo, Y., Zhu, L.-Z., Wan, Z.-Y., and Lu, B.-L. (2020). Data augmentation for enhancing eeg-based emotion recognition with deep generative models. *J. Neural Eng.* 17:056021. doi: 10.1088/1741-2552/abb580

Maas, A. L., Hannun, A. Y., Ng, A. Y., et al. (2013). "Rectifier nonlinearities improve neural network acoustic models," in *Proceedings of ICML, Vol. 30* (Atlanta), 3.

Marafioti, A., Perraudin, N., Holighaus, N., and Majdak, P. (2019). "Adversarial generation of time-frequency features with application in audio synthesis," in *Proceedings of the 36th International Conference on Machine Learning, ICML 2019*, eds K. Chaudhuri and R. Salakhutdinov (Long Beach, CA), 4352–4362.

Minecan, D., Natarajan, A., Marzec, M., and Malow, B. (2002). Relationship of epileptic seizures to sleep stage and sleep depth. *Sleep* 25, 56–61. doi: 10.1093/sleep/25.8.56

Mirza, M., and Osindero, S. (2014). Conditional generative adversarial nets. *CoRR, abs/1411.1784.* doi: 10.48550/arXiv.1411.1784

Odena, A., Dumoulin, V., and Olah, C. (2016). Deconvolution and checkerboard artifacts. *Distill* 1:e3. doi: 10.23915/distill.00003

Plummer, C., Harvey, A. S., and Cook, M. J. (2008). Eeg source localization in focal epilepsy: where are we now? *Epilepsia* 49, 201–218. doi: 10.1111/j.1528-1167.2007.0138

Radford, A., Metz, L., and Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434.* doi: 10.48550/arXiv.1511.06434

Ramantani, G., Maillard, L., and Koessler, L. (2016). Correlation of invasive EEG and scalp EEG. *Seizure* 41, 196–200. doi: 10.1016/j.seizure.2016.05.018

Salimans, T., Goodfellow, I. J., Zaremba, W., Cheung, V., Radford, A., and Chen, X. (2016). "Improved techniques for training GANs," in *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016*, eds D. D. Lee, M. Sugiyama, U. von Luxburg, I. Guyon, and R. Garnett (Barcelona), 2226–2234.

Schirrmeister, R. T., Gemein, L., Eggensperger, K., Hutter, F., and Ball, T. (2017). Deep learning with convolutional neural networks for decoding and visualization of EEG pathology. *In 2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB).* 1–7. doi: 10.1109/SPMB.2017.8257015

Sejdic, E., Djurovic, I., and Stankovic, L. (2008). Quantitative performance analysis of scalogram as instantaneous frequency estimator. *IEEE Trans. Signal Process.* 56, 3837–3845. doi: 10.1109/TSP.2008.924856

Selim, M., Zhang, J., Fei, B., Zhang, G.-Q., and Chen, J. (2020). "Stan-CT: Standardizing CT image using generative adversarial networks," in *AMIA Annual Symposium Proceedings, Vol. 2020* (American Medical Informatics Association), 1100.

Smith, K. E., and Smith, A. O. (2020). Conditional gan for timeseries generation. *arXiv preprint arXiv:2006.16477.* doi: 10.48550/arXiv.2006.16477

Smith, K. E., and Smith, A. O. (2021). A spectral enabled GAN for time series data generation. *arXiv preprint arXiv:2103.01904.* doi: 10.48550/arXiv.2103.01904

Spyrou, L., and Sanei, S. (2016). "Coupled dictionary learning for multimodal data: an application to concurrent intracranial and scalp EEG," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Shanghai), 2349–2353. doi: 10.1109/ICASSP.2016.7472097

Tatum, W. O. (2014). Ellen R. grass lecture: extraordinary EEG. *Neurodiagn. J.* 54, 3–21. doi: 10.1080/21646821.2014.11079932

Wang, J., Liang, S., He, D., Wang, Y., Wu, Y., and Zhang, Y. (2020). "A sequential graph convolutional network with frequency-domain complex network of EEG signals for epilepsy detection," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (Seoul), 785–792. doi: 10.1109/BIBM49941.2020.9313232

Yao, Y., Plested, J., and Gedeon, T. (2018). "Deep feature learning and visualization for EEG recording using autoencoders," in *International Conference on Neural Information Processing* (Siem Reap: Springer), 554–566. doi: 10.1007/978-3-030-04239-4_50

Yao, Y., Plested, J., and Gedeon, T. (2020). Information-preserving feature filter for short-term EEG signals. *Neurocomputing* 408, 91–99. doi: 10.1016/j.neucom.2019.11.106

Yue, Z., Wang, Y., Duan, J., Yang, T., Huang, C., Tong, Y., et al. (2022). "TS2VEC: towards universal representation of time series," in *Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 36*, 8980–8987. doi: 10.1609/aaai.v36i8.20881

Zhang, X., Zhao, Z., Tsiligkaridis, T., and Zitnik, M. (2022). Self-supervised contrastive pre-training for time series via time-frequency consistency. *arXiv preprint arXiv:2206.08496.* doi: 10.48550/arXiv.2206.08496

Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 08 September 2022
DOI 10.3389/fnins.2022.1000716

Emotion recognition based on multi-modal physiological signals and transfer learning

Zhongzheng Fu[†], Boning Zhang[†], Xinrun He*, Yixuan Li, Haoyuan Wang and Jian Huang

School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China

EDITED BY
Xiaomao Fan,
South China Normal University, China

REVIEWED BY
Liping Xie,
Wuhan University of Technology,
China
Yanling Hao,
Queen Mary University of London,
United Kingdom

*CORRESPONDENCE
Xinrun He
hexr@hust.edu.cn

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

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

RECEIVED 22 July 2022
ACCEPTED 18 August 2022
PUBLISHED 08 September 2022

CITATION
Fu Z, Zhang B, He X, Li Y, Wang H and Huang J (2022) Emotion recognition based on multi-modal physiological signals and transfer learning. *Front. Neurosci.* 16:1000716. doi: 10.3389/fnins.2022.1000716

In emotion recognition based on physiological signals, collecting enough labeled data of a single subject for training is time-consuming and expensive. The physiological signals' individual differences and the inherent noise will significantly affect emotion recognition accuracy. To overcome the difference in subject physiological signals, we propose a joint probability domain adaptation with the bi-projection matrix algorithm (JPDA-BPM). The bi-projection matrix method fully considers the source and target domain's different feature distributions. It can better project the source and target domains into the feature space, thereby increasing the algorithm's performance. We propose a substructure-based joint probability domain adaptation algorithm (SSJPDA) to overcome physiological signals' noise effect. This method can avoid the shortcomings that the domain level matching is too rough and the sample level matching is susceptible to noise. In order to verify the effectiveness of the proposed transfer learning algorithm in emotion recognition based on physiological signals, we verified it on the database for emotion analysis using physiological signals (DEAP dataset). The experimental results show that the average recognition accuracy of the proposed SSJPDA-BPM algorithm in the multimodal fusion physiological data from the DEAP dataset is 63.6 and 64.4% in valence and arousal, respectively. Compared with joint probability domain adaptation (JPDA), the performance of valence and arousal recognition accuracy increased by 17.6 and 13.4%, respectively.

KEYWORDS

emotion recognition, transfer learning, domain adaptation, physiological signal, multimodal fusion, individual difference

Introduction

Emotion is a complex expression that integrates people's psychological and physiological functions. It reflects the subjective response of individuals to external stimuli all the time (Sharot et al., 2004). Since affective computing was proposed, researchers have devoted to digitizing the concept of emotion, enabling computers to recognize and process it, and providing more reliable signal input for human-computer interaction (Picard, 2003; Mühl et al., 2014). In the human-computer interaction system, accurately decoding the user's emotion can make the device not only passively receive the user's instructions but also truly perceive the user's state, to better understand the user's intention and establish a more natural and harmonious human-computer interaction environment (Egger et al., 2019). As a research hotspot in human-computer interaction, affective computing is widely used in traffic safety (Liu et al., 2019; Du et al., 2020), brain-computer interface (Al-Nafjan et al., 2017; Rao et al., 2018), medical health (Hosseinifard et al., 2013; Huang et al., 2019), and other fields. Affective computing includes three continuous processes: emotion recognition, behavior generation, and induction. Accurate emotion recognition is the basis for building a good human-computer interaction experience (Egger et al., 2019). However, in practical applications, collecting large numbers of data for each user to train the classifier is difficult, and the recognition accuracy is easily affected by data noise (Wan et al., 2021). When the accuracy of emotion recognition is influenced by physiological signals' individual differences and inherent noise, making the model trained in the existing data set accurately identify new users' emotions without collecting data or collecting as little data as possible has essential research value and application significance.

Nowadays, there are many emotion recognition methods, such as analyzing users' voices (Li et al., 2019; Shaqra et al., 2019), facial expressions (Lawrence et al., 2015; Abdulsalam et al., 2019), and physiological signals (He et al., 2017; Liao et al., 2020). Physiological signals are the most easily acquired signals by the human body through sensors. It contains many important physiological and psychological information about the human body and plays a significant role in computer recognition of human emotions (Li et al., 2021). Compared with emotion recognition based on facial expression, emotion recognition based on physiological signals not only has the advantages of low cost and high efficiency in data acquisition but also can avoid the errors caused by light and shadow acquisition and the invasion of user privacy (Hao et al., 2020; Fu et al., 2021).

In the aspect of emotion recognition, electroencephalogram (EEG) has been paid more attention by researchers among many physiological signals. The analysis of EEG signals in the field of emotion recognition depends on data preprocessing, feature extraction, and feature classification (Xie et al., 2021). Many researchers use traditional machine learning or deep neural network to classify EEG signals by extracting the energy features of the delta, theta, alpha, beta, and gamma bands. For example, Verma and Tiwary (2014) extracted the relative power energy, logarithmic relative power energy, absolute logarithmic relative power energy, standard deviation, and spectral entropy features of five frequency bands from EEG signals. Liu et al. (2016) used a deep autoencoder to extract the features of EEG signals in the DEAP dataset and extract features. Sorkhabi (2014) used continuous wavelet transform to extract energy features of five frequency bands and entropy features of wavelet coefficients. Yin et al. (2017) extracted the frequency band power features, statistical features, signal zero crossing rate, Shannon entropy, spectral entropy, kurtosis, skewness, and other features of the five frequency bands. Torres-Valencia et al. (2017) extracted statistical features of EEG signals and power features of five frequency bands.

However, a single EEG signal's lack of feature information will lead to low emotion recognition accuracy. Some researchers use feature level fusion or signal level fusion to fuse multimodal signals to improve emotion recognition accuracy. He et al. (2021) extracted 11 features from the EEG signal of FP2 channel and 6 features from HR. Using multi-core learning for fusion, they achieved 67% binary classification accuracy under fewer signals and channels. Song et al. (2019) used attention-based long-term short-term memory to fuse multimodal physiological signals, including electroencephalogram (EEG), Galvanic Skin Response (GSR), respiration (RSP), and electrocardiogram (ECG), to improve the classification accuracy. Our study uses EEG, RSP, GSR, and photo-plethysmograph (PPG) signals collected from the database for emotion analysis using physiological signals (DEAP) (Koelstra et al., 2011) for feature extraction. We concatenate four mode features to achieve multimodal feature fusion. It can remedy the inherent limitations of the single mode by providing more dimensional features. Consequently, the multimodal features improve the accuracy of emotion recognition.

The EEG signals are very complex due to the inherent non-stationary, non-linear, and non-Gaussian characteristics (Subha et al., 2010). Meanwhile, EEG signals are greatly affected by age, psychology, and other factors, which result in significant differences in individual EEG signals (Lotte et al., 2018). This difference is often substantial and cannot be ignored. The traditional emotion recognition based on EEG does not consider the existence of differences and directly trains a general model. The difference between EEG signals of different individuals will directly affect the accuracy of model recognition and classification and lead to a poor generalization ability of the model (Zheng and Lu, 2016). Considering different types of information in EEG signals make it difficult to filter out information sensitive to specific tasks, and there are few similar EEG data among different individuals due to the significant difference in EEG, it is problematic that use the deep learning model based on old user data training to estimate the mental state of new users (Wan et al., 2021).

In order to solve the above problems, some researchers have introduced transfer learning to emotion recognition. Li et al. (2019) proposed a multi-source transfer learning algorithm to transfer the existing emotion model to new subjects. The experimental results show that this method can effectively reduce the demand for data quantity and increase the calibration capability of the model. Chai et al. (2017) proposed an adaptive subspace feature matching algorithm for emotion recognition, which aligns the source and target subspaces by learning linear transformation to reduce the distribution discrepancy between the source and target domains. Lin and Jung (2017) proposed a conditional transfer learning framework. The algorithm first evaluates the individual's transferability to positive transfer and then selectively leverages the data from others with comparable feature spaces. Therefore, in order to solve the low accuracy of emotion recognition caused by the mismatch between individual specificity and global threshold, we introduce domain adaptation, a transfer learning method, into emotion recognition. This method can apply the patterns learned in one domain to other domains and reduce the differences in EEG data distribution so that to improve the model's ability to recognize new users' emotions.

Generally, the domain adaptation method usually seeks the alignment between the source and target domains. Different domain adaptation methods often use different alignments. The current alignment methods can be divided into three categories according to distribution matching schemes: domain-level, class-level, and sample-level (Lu et al., 2021). Pan et al. (2010) proposed the transfer component analysis (TCA) method, which uses the maximum mean difference (MMD) to learn a transformation matrix in the reproducing kernel Hilbert space (RKHS) to align the marginal distribution between the two domains. Long et al. (2017) proposed the joint distribution adaptation (JDA) method to align the joint distribution of multiple domains through multi-kernel MMD. Sun et al. (2017) proposed the correlation alignment (CORAL) method, which minimizes the domain shift by aligning the second-order statistical data of source and target distribution. The above commonly used domain adaptation methods belong to domain-level matching. The domain-level matching completely ignores the intra-domain data structure. It is too rough to miss some details and challenging to achieve good matching results.

The sample level matching can avoid the problem that domain-level matching ignores intra-domain data structure. Courty et al. (2017) proposed a regularized unsupervised optimal transport model, which uses the optimal transport theory to calculate the distance between the probability distributions of the source and the target domain. In the research of Das and Lee (2018), the source and the target domain are regarded as hypergraphs, and the first-order, second-order, and third-order similarities between graphs are used for class-regularized hypergraph matching to obtain the matching between the samples of the source domain and the target domain. However, sample level matching is very time-consuming, and it is more prone to overfitting when local information is affected by the noise.

Class-level matching can neutralize too rough domain-level matching and too fine sample-level matching. Wang et al. (2018) proposed the Stratified Transfer Learning (STL) method. STL transforms the same classes in the source and the target domain into the same subspace and uses the intra-affinity of the class to perform knowledge migration within the class. Tian et al. (2020) proposed the Centroid Matching and local Manifold Self-learning (CMMS) method. CMMS can thoroughly explore the data distribution structure of the domain and minimize the distribution difference in domain adaptation by combining class centroid matching with local manifold self-learning. Lu et al. (2021) proposed a domain adaptation method based on substructure level matching, which regards a class as synthesizing multiple substructures and aligning the substructures. The above commonly used domain adaptation methods belong to class-level matching. Considering that the EEG signal acquisition process contains the location information of different channels, which has the intra-domain data structure, we adopt the class-level domain adaptation to avoid rough alignment of domain-level adaptation and overfitting of sample-level adaptation.

In the matching process of the source and target domains, it is necessary to project the source and target domains into the same feature space through the projection matrix. The TCA, JDA, BDA, and JPDA all uses the single projection matrix for transfer (Pan et al., 2010; Long et al., 2017; Wang et al., 2017; Zhang et al., 2020). However, the distribution of the source domain and target domain is different, and a single projection matrix cannot account for all the feature distribution of the source and target domains. Therefore, we propose a bi-projection matrix (BPM) to better project the source and target domains into the feature space.

This paper uses EEG, RES, PPG, and GSR signals collected from the DEAP dataset to extract features, and we concatenate four mode features to achieve multimodal feature fusion. Multimodal fusion gives full play to the advantages of each mode and makes up for its inherent limitations, improving the accuracy of emotion recognition. In order to improve the generalization ability of the model, we propose a joint probability domain adaptation method based on the substructure. Substructure-level data is aligned by discriminative joint probability maximum mean discrepancy (DJP-MMD) (Zhang et al., 2020). Substructure-based joint probability domain adaptation (SSJPDA) can avoid inadaptability caused by rough matching and overfitting when learning local information caused by noise points. In order to better project the source and target domains, we propose a method of the bi-projection matrix (BMP), which can effectively avoid data loss in the projection stage.

The main contributions of this study are as follows:

- We proposed a substructure-based joint probability domain adaptation method (SSJPDA).
- We proposed the bi-projection matrix (BPM) method and applied it to the SSJPDA algorithm.
- We validated the SSJPDA algorithm and the SSJPDA-BPM method on DEAP dataset.

The rest of this paper is arranged as follows: Section "Materials and methods" introduces the SSJPDA with the BPM algorithm. Section "Results" presents the results verified on the DEAP dataset. Section "Discussion" gives the full discussion above the result.

Materials and methods

Physiological signal dataset

This study adopted the DEAP dataset to inspect our proposed algorithm. DEAP dataset was established by Koelstra et al. (2011) in 2012 and contained 32 subjects. Every subject watched the 40 selected music videos, and each video viewed by the subjects was regarded as an independent experiment. After the video viewing, the subjects need to use the self-evaluation model to score arousal, valence, like/dislike, dominance, and familiarity, providing label information for each signal. Every experiment recorded 40 physiological signals of subjects, of which the first 32 signals were EEG signals collected according to the international 10–20 system, and the remaining 8 signals were peripheral physiological signals, including 2 ophthalmic signals, 1 skin electrical signal, 2 EMG signals, 1 respiratory record, 1 plethysmography, and 1 temperature record. The dataset also preprocessed the collected signals. Each test section's EEG data and other peripheral physiological signal data were divided into 3 s baseline data and 60 s test data. EEG signals are collected according to the international 10–20 lead system and down-sampling from 512 Hz original sampling frequency to 128 Hz. RES, PPG, and GSR signals are down-sampled to 128 Hz. A band-pass frequency filter of 4–45 Hz and a blind source separation technique were used to remove the eye artifacts.

Feature extraction

Considering that the subjects are not always in a high emotional activation state if the sliding window is used to divide the data into small segments, many segments will contain useless information (Piho and Tjahjadi, 2018). Therefore, we directly extract features from the preprocessed 60 s experimental data to make samples instead of dividing continuous data into multiple segments and making each segment into samples in the feature processing. We extract the differential entropy features of five frequency bands from each recorded EEG data from each EEG channel. These five frequency bands are related to people's state of mind, so they also contain information about the state of specific thinking tasks. These five bands are Delta (1–4 Hz), Theta (4–8 Hz), Alpha (8–13 Hz), Beta (13–30 Hz), and Gamma (30–48 Hz). Some studies have shown that the differential entropy feature is superior to the power spectral density feature (PSD) in EEG-based emotion recognition (Zheng and Lu, 2015; Soleymani et al., 2017).

We extract their time-domain and frequency-domain features for the peripheral physiological signals PPG, GSR, and RES. The extracted time-domain features and frequency-domain features refer to numerous previous studies (Verma and Tiwary, 2014; Yin et al., 2017; Zhang et al., 2021). Time-domain features depend on statistical features, which are simple and intuitive. It realizes classification by analyzing statistical features such as mean, maximum, minimum, root mean square, standard deviation, etc. The time-domain analysis contains all the characteristics of physiological signals, and the signal is processed directly. Hence the loss of information is relatively small. For example, from the time domain characteristics of PPG signals, we can analyze the heart rate and its changes, which are closely related to emotional arousal. In addition, Frequency domain features can show the frequency information that time-domain features cannot reach in more detail. Consequently, we got 1,280 samples (32 subjects × 40 samples). **Table 1** lists the features extracted from the data.

The generation of substructures

The source domain $\{X_S, Y_S\} = \{(\mathrm{x}_{s,i}, y_{s,i})\}_{i=1}^{n_s}$containing the label recorded as $\mathcal{D}_s$. The target domain $X_t = \{\mathrm{x}_{t,j}\}_{j=1}^{n_t}$ without label recorded as$\mathcal{D}_t$. The n_s and n_t are the number of source domain samples and target samples, respectively. $\mathrm{x} \in \mathbb{R}^{d\times 1}$is the feature vector, and $y \in \{1, \ldots, C\}$ is its label in the C-class classification problem. $\mathcal{D}_s$ and $\mathcal{D}_t$ have the same feature space and label space, but the feature distribution is different, i.e., $P(X_s, Y_s) \neq P(X_t, Y_t)$. The task of domain adaptation is to reduce the distribution difference between the source domain and the target domain, so as to predict the label $y_t$of the target domain $\mathcal{D}_t$ with the help of the source domain $\mathcal{D}_s$ (Lu et al., 2021).

We use $\delta \sim \mathcal{N}\left(0; \sigma^2\right)$ and X to represent all feature data and the Gaussian mixture model (GMM) to fit them. The kth component in GMM is recorded as $X_k \sim \mathcal{N}(z_k, \sigma_k)$ where z_k represents mean value and σ_k represents covariance. Our goal is to get mean value z_k and covariance σ_k. These GMM parameters can be obtained using the Expectation Maximum (EM) algorithm. Suppose K_s and K_t are the number of GMM

TABLE 1 The features used in Experiment 1 and Experiment 2.

Signal	Feature	Description	Dimension
EEG	Differential Entropy (DE)	DE in different bands: Delta (1–4 Hz), Theta (4–8 Hz), Alpha (8–13 Hz), Beta (13–30 Hz), and Gamma (30–48 Hz)	32 channels × 5 features
PPG	Time Domain	Mean value, maximum value, minimum value, standard deviation and root mean square value of heart rate interval. Heart rate (times/second)	1 channel × 8 features
	Frequency domain	Power spectral density of bands 0.1–1.5 Hz and 1.5–3 Hz.	
GSR	Time Domain	Mean, standard deviation	1 channel × 7 features
	Frequency domain	Power spectral density of bands 0.4–0.8 Hz, 0.8–1.2 Hz, 1.2–1.6 Hz, 1.6–2.0 Hz, and 2.0–2.4 Hz.	
RES	Time Domain	Mean, maximum, minimum, standard deviation, and root mean square value of respiratory interval. Respiratory rate (times / second)	1 channel × 8 features
	Frequency domain	Power spectral density of bands 0.1–1.5 Hz and 1.5–3.0 Hz.	

components in the source domain and the target domain, respectively. K_s is determined by the Bayesian Information Criterion (BIC), and K_t is manually set according to the specific data set.

After obtaining the GMM of the source domain and the target domain, we regard each component of the GMM as a substructure in the feature space, and the information of the cluster center represents the substructure. Specifically, set

$$\mu_s = \sum_{i=1}^{k_s} w_{s,i}\delta_{z_{s,i}} \tag{1}$$

$$\mu_t = \sum_{i=1}^{k_t} w_{t,i}\delta_{z_{t,i}} \tag{2}$$

where μ_s and $\mu_t$are the distribution of source domain and target domain, respectively. $z \in \mathbb{R}^{d\times 1}$ is cluster center, and δ_z is the Dirac function at location z. $w$is the probability weight associated with z, where $\sum_{i=1}^{k_s} w_{s,i} = 1$ and $\sum_{i=1}^{k_t} w_{t,i} = 1$.

The cost between z_i and z_j in square Euclidean distance can be expressed as

$$c\left(z_{s,i}, z_{t,j}\right) = \left|\left|z_{s,i} - z_{t,j}\right|\right|_2^2 \tag{3}$$

Therefore, the problem can be regarded as the partial optimal transmission (POT) problem, and the upper bound $w_{s,i}$is 1. The total cost of POT is $\langle\pi, C\rangle_F$ that is the Frobenius dot product of cost matrix C and coupling matrix π. The $C \in \mathbb{R}^{k_s\times k_t}$ represents the cost of μ_s and $\mu_t$distribution, and the$\pi \in \mathbb{R}^{k_s\times k_t}$ represents the coupling between μ_s and μ_t distribution.

The goal is to obtain the optimal transmission, which can be expressed as

$$\begin{aligned} &\pi_1^* = \arg\min_\pi \langle\pi, C\rangle_F + \lambda_1 H(\pi) \\ &s.t.\pi^T 1_{k_s} = w_t \end{aligned} \tag{4}$$

where $H(\pi) = \sum_{ij} \pi_{ij}\log\pi_{ij}$ is the entropy term, and λ_1 is the super parameter to balance the speed and accuracy calculation.

The feasible solution set of $\pi^T 1_{k_s} = w_t$ is C_1, and then it can be solved by the Lagrange method. Thus, we can easily get the optimal π^*.

$$\pi_1^* = \pi_0 diag\left(w_t \oslash \pi_0^T 1_{k_s}\right) \tag{5}$$

where $\pi_0 = \exp\left(-\frac{C}{\lambda_1} - 1\right)$ and $\oslash$ represent element-wise divide and $diag$represents the diagonals. Once the coupling matrix π_1^* is obtained, the source domain weights can be easily calculated as $w_s = \pi_1^*\, 1_{k_t}$.

Substructural joint probability domain adaptation

The domain adaptation (DA) method attempts to find a mapping h. The source domain and target domain are mapped to the same subspace, so that the classifier trained on $h(x_s)$can achieve good classification effect on $h(x_t)$. For example, a linear map$h(x) = A^T x$for the source and the target domains, where $A \in \mathbb{R}^{d\times p}, p \leq d$.

Due to the difference between the source domain and the target domain, it is generally assumed that their probabilities distributions are not equal. The derivation of TCA, JDA and BDA algorithms are based on the inequality of the marginal probabilities $P(X_s) \neq P(X_t)$ or the conditional probabilities $P(Y_s|X_s) \neq P(Y_t|X_t)$. However, the JPDA algorithm derives from the inequality assumption of joint probabilities $P(X_s, Y_s) \neq P(X_t, Y_t)$. Because JPDA directly considers the difference of joint probability distribution, the performance of JPDA is better than the traditional DA method, which JPDA can improve the between-domain transferability and the between-class discrimination (Zhang et al., 2020).

After obtaining the substructure, the set of substructures in source domain is recorded as $\{Z_S, Y_S'\} = \{(z_{s,i}, y_{s,i}')\}_{i=1}^{k_s}$, and the set of substructures in target domain is recorded as $Z_t = \{z_{t,j}\}_{j=1}^{k_t}$, where k_s and k_t are the number of source domain substructure and target domain substructure, respectively.

Let the source domain substructure one-hot coding label matrix be $Y_s' = [y_{s,1}'; \ldots; y_{s,k_s}']$ and the predicted target domain substructure one-hot coding label matrix be $\hat{Y}_t' =$

$[\hat{y}'_{t,1}; \ldots ; \hat{y}'_{t,k_t}]$ where $y'_{s,ks} \in \mathbb{R}^{1\times C}$ and $\hat{y}'_{t,kt} \in \mathbb{R}^{1\times C}$. Define

$$F_s = [Y'_s(:, 1)* (C-1) , ..., Y'_s(:, C) * (C-1)] \tag{6}$$

$$\hat{F}_t = [\hat{Y}'_t(:, 1 : C)_{\hat{c}\neq 1}, ..., \hat{Y}'_t(:, 1 : C)_{\hat{c}\neq C}] \tag{7}$$

where $Y'_s(:, c)$ denotes the c-th column of Y'_s, $Y'_s(:, c) * (C-1)$ repeats $Y'_s(:, C)$ $C-1$ times to form a matrix in $\mathbb{R}^{k_s\times(C-1)}$, and $\hat{Y}'_t(:, 1 : C)_{\hat{c}\neq c}$ is formed by the 1st to the C-th, (except the c-th) columns of Y'_t. Clearly, $F_s \in \mathbb{R}^{k_s\times(C(C-1))}$ and $\hat{F}_t \in \mathbb{R}^{k_t\times(C(C-1))}$. F_s is fixed, and $\hat{F}_t$ is constructed from the pseudo labels, which are updated iteratively.

Therefore, the objective function of JPDA can be written as follows:

$$\begin{aligned}&\min_A ||A^{\mathrm{T}}Z_sN_s - A^{\mathrm{T}}Z_tN_t||_F^2 - \mu||A^{\mathrm{T}}Z_sM_s - A^{\mathrm{T}}Z_tM_t||_F^2 + \lambda||A||_F^2\\ &\mathrm{s.}t.A^{\mathrm{T}}ZHZ^{\mathrm{T}}A = I\end{aligned} \tag{8}$$

where $\mu > 0$ is a trade-off parameter and λ is a regularization parameter. N_s, N_t, M_s and M_t are defined as

$$N_s = \frac{Y'_s}{k_s}, N_t = \frac{\hat{Y}'_t}{k_t} \tag{9}$$

$$M_s = \frac{F_s}{k_s}, M_t = \frac{\hat{F}_t}{k_t} \tag{10}$$

where $H = I - 1_k$ is the centering matrix, in which $k = k_s + k_t$ and $1_k \in \mathbb{R}^{k\times k}$ is a matrix with all elements being $\frac{1}{k}$.

Let $Z = [Z_s, Z_t]$, then we reach the Lagrange function of Eq. 8

$$\mathcal{J} = \mathrm{tr}(A^{\mathrm{T}}(Z(R_{\min} - \mu R_{\max})Z^{\mathrm{T}} + \lambda I)A)+\mathrm{tr}(\eta(I - A^{\mathrm{T}}ZHZ^{\mathrm{T}}A)) \tag{11}$$

where η is Lagrange multiplier, and

$$R_{\min} = \begin{bmatrix} N_sN_s^{\mathrm{T}} & -N_sN_t^{\mathrm{T}} \\ -N_tN_s^{\mathrm{T}} & N_tN_t^{\mathrm{T}} \end{bmatrix} \tag{12}$$

$$R_{\max} = \begin{bmatrix} M_sM_s^{\mathrm{T}} & -M_sM_t^{\mathrm{T}} \\ -M_tM_s^{\mathrm{T}} & M_tM_t^{\mathrm{T}} \end{bmatrix} \tag{13}$$

$R_{\max}$ and $R_{\min}$ have dimensionality $k \times k$.

By setting the derivative $\nabla_A\mathcal{J} = 0$, Eq. 17 becomes a generalized eigen-decomposition problem:

$$(Z(R_{\min} - \mu R_{\max})Z^{\mathrm{T}} + \lambda I)A{=}\eta ZHZ^{\mathrm{T}}A \tag{14}$$

A is then formed by the p trailing eigen-vectors. A classifier can then be trained on$A^{\mathrm{T}}Z_s$ and applied to $A^{\mathrm{T}}Z_t$.

The pseudocode of SSJPDA for classification is summarized in Algorithm 1.

```
Input:
  $X_S$ and $X_t$, source and target domain
  feature matrices;
  $Y_S$, source domain one-hot coding label
  matrix;
  $\mu$, trade-off parameter;
  $\lambda$, regularization parameter;
  $T$, number of iterations;
Output:
  $\hat{Y}_t$, estimated target domain labels.
Begin:
  Use EM for GMM, cluster each class
  data in the source to obtain
  $\{Z_S, Y'_S\} = \{(z_{s,i}, y'_{s,i})\}_{i=1}^{k_s}$, and cluster the
  unlabeled data in target domain to
  obtain $Z_t = \{z_{t,j}\}_{j=1}^{k_t}$;
  Compute cost matrix $C$ and coupling
  matrix $\pi$ using Eq. 3 and Eq. 4,
  respectively;
  Compute the weights of source
  substructures $w_s = \pi_1^* 1_{k_t}$ and target
  substructures $w_t = \frac{1_{k_t}}{k_t}$
  for $n$ = 1, . . . , T do
    Construct the joint probability
    matrix $R_{\min}$ and $R_{\max}$ by Eq. 12 and
    Eq. 13;
    Solve the generalized
    eigen-decomposition problem in
    Eq. 14 and select the $p$ trailing
    eigenvectors to construct the
    projection matrix A;
    Train a classifier $f$ on $(A^{\mathrm{T}}Z_s, Y'_S)$ and
    apply it to $A^{\mathrm{T}}Z_t$ to obtain
    $\hat{Y}'_t = \{y'_{t,j}\}_{j=1}^{k_t}$ which is the label matrix
    of substructure in target domain
    $Z_t = \{z_{t,j}\}_{j=1}^{k_t}$
  End for
  For each substructure $z_{t,j}$, assign its
  label $y'_{t,j}$ to all samples it contains,
  and gets$\hat{Y}_t = \{y_{t,j}\}_{j=1}^{n_t}$
End
```

Algorithm 1. Substructural Joint Probability Distribution Adaptation (SSJPDA)

Substructure-based joint probability domain adaptation algorithm with bi-projection matrix

As described in the previous subsection, the source and target domains have different probability distributions, so applying only a single projection matrix to both domains simultaneously may lack the ability to align their probability distributions well. It is better to make the source domain and the target domain have their own projection matrix to accomplish the distribution alignment task together. On this basis, we take SSJPDA algorithm as an example to explain how to design the projection matrix of source domain and target domain, respectively, and call it SSJPDA-BPM.

Donate the projection matrices of the source domain and the target domain asA_s and A_t, respectively. Therefore, the objective function of SSJPDA-BPM can be written as follows:

$$\begin{aligned} &\min_{A} ||A_s^{\mathrm{T}} Z_s N_s - A_t^{\mathrm{T}} Z_t N_t||_F^2 - \mu ||A_s^{\mathrm{T}} Z_s M_s - A_t^{\mathrm{T}} Z_t M_t||_F^2 \\ &\qquad\qquad + \lambda \left(||A_s||_F^2 + ||A_t||_F^2 \right) \\ &s.t. A_s^{\mathrm{T}} Z_s H_s Z_s^{\mathrm{T}} A_s = I_{k_s}, A_t^{\mathrm{T}} Z_t H_t Z_t^{\mathrm{T}} A_t = I_{k_t} \end{aligned} \tag{15}$$

where $H_s = I_{k_s} - 1_{k_s}$ (or $H_t = I_{k_t} - 1_{k_t}$) is the centering matrix, in which $1_{k_s} \in \mathbb{R}^{k_s \times k_s}$ (or $1_{k_t} \in \mathbb{R}^{k_t \times k_t}$) is a matrix with all elements being $\frac{1}{k_s}$ (or $\frac{1}{k_t}$).

Let$Z_A = [A_s^{\mathrm{T}} Z_s, A_t^{\mathrm{T}} Z_t]$, then we reach the Lagrange function of Eq. 15

$$\begin{aligned} \mathcal{J} &= tr(Z_A R Z_A^{\mathrm{T}}) + tr(\eta_s (I_{k_s} - A_s^{\mathrm{T}} Z_s H_s Z_s^{\mathrm{T}} A_s)) \\ &+ tr(\eta_t (I_{k_t} - A_t^{\mathrm{T}} Z_t H_t Z_t^{\mathrm{T}} A_t)) + tr(A_s^{\mathrm{T}} A_s) + tr(A_t^{\mathrm{T}} A_t) \end{aligned} \tag{16}$$

where η_s η_t are Lagrange multipliers, and

$$\begin{aligned} R = \ R_{\min} - \mu R_{\max} &= \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} \\ &= \begin{bmatrix} N_s N_s^{\mathrm{T}} - \mu M_s M_s^{\mathrm{T}} & -N_s N_t^{\mathrm{T}} + \mu M_s M_t^{\mathrm{T}} \\ -N_t N_s^{\mathrm{T}} + \mu M_t M_s^{\mathrm{T}} & N_t N_t^{\mathrm{T}} - \mu M_t M_t^{\mathrm{T}} \end{bmatrix} \end{aligned} \tag{17}$$

By setting the derivative $\nabla_{A_s} \mathcal{J} = 0$, $\nabla_{A_s} \mathcal{J} = 0$, and add a constraint $Z_s R_{12} Z_t^T A_s {=} Z_t R_{21} Z_s^T A_t$, then Eq. 16 becomes two generalized eigen-decomposition problem:

$$(Z_s R_{11} Z_s^{\mathrm{T}} + Z_t R_{21} Z_s^{\mathrm{T}} + \lambda I) A_s {=} \eta_s Z_s H_s Z_s^{\mathrm{T}} A_s \tag{18}$$

$$(Z_t R_{22} Z_t^{\mathrm{T}} + Z_s R_{12} Z_t^{\mathrm{T}} + \lambda I) A_t {=} \eta_t Z_t H_t Z_t^{\mathrm{T}} A_t \tag{19}$$

A_s and A_t are then formed by the p trailing eigen-vectors of each problem. A classifier can then be trained on $A_s^{\mathrm{T}} Z_s$ and applied to ${A_t}^{\mathrm{T}} Z_t$.

The pseudocode of SSJPDA-BPM for classification is summarized in **Algorithm 2**.

```
Input:
  $X_S$ and $X_t$, source and target domain feature matrices;
  $Y_S$, source domain one-hot coding label matrix;
  $\mu$, trade-off parameter;
  $\lambda$, regularization parameter;
  $T$, number of iterations;
Output:
  $\hat{Y}_t$, estimated target domain labels.
Begin:
  Use EM for GMM, cluster each class data in the source to obtain $\{Z_S, Y'_S\} = \{(z_{s,i}, y'_{s,i})\}_{i=1}^{k_s}$, and cluster the unlabeled data in target domain to obtain $Z_t = \{z_{t,j}\}_{j=1}^{k_t}$;
  Compute cost matrix $C$ and coupling matrix $\pi$ using Eq. 3 and Eq. 4 respectively;
  Compute the weights of source substructures $w_s = \pi_1^* 1_{k_t}$ and target substructures $w_t = \frac{1_{k_t}}{k_t}$
  for n = 1,..., T do
    Construct the joint probability matrix $R$ in Eq. 17
    Solve the generalized eigen-decomposition problem in Eq. 18 and Eq. 19, and select the $p$ trailing eigenvectors to construct the projection matrix $A_s$ and $A_t$;
    Train a classifier $f$ on $A_s^{\mathrm{T}} Z_s$ and applied to $A_t^{\mathrm{T}} Z_t$ to obtain $\hat{Y}'_t = \{y'_{t,j}\}_{j=1}^{k_t}$ which is the label matrix of substructure in target domain $Z_t = \{z_{t,j}\}_{j=1}^{k_t}$
  End for
  For each substructure $z_{t,j}$, assign its label $y'_{t,j}$ to all samples it contains, and gets$\hat{Y}_t = \{y_{t,j}\}_{j=1}^{n_t}$
End
```

Algorithm 2. Substructural Joint Probability Distribution Adaptation with Bi-Projection Metrix (SSJPDA-BPM)

Validation of the substructure-based joint probability domain adaptation algorithm and substructural joint probability distribution adaptation with bi-projection metrix

The DEAP dataset contains 32 subjects, each taking turns as the target domain and the remaining 31 people as the source domain. The number of samples in the source domain is 1,240 (31 subjects × 40 samples), and the number of target domain samples is 40 (1 subject × 40 samples). After dividing the source and target domains, the EEG, GSR, PPG, and RES modes were transferred, respectively, and all the subjects' valence and arousal dimensions were classified, respectively. In each sample, the feature dimension of EEG is 160, the feature dimension of GSR is 7, the feature dimension of PPG is 8, and the feature dimension of RES is 8. Those four modes were fused through average splicing, where the feature dimension after fusion in each sample is 183. The feature dimension of the modes remains the same dimension before and after the transfer learning. The effects of single-mode transfer and multi-mode transfer are compared to explore whether data fusion can promote the accuracy of the transfer learning algorithm. By comparing SSJPDA with other transfer learning methods and traditional machine learning methods, this paper explores whether SSJPDA can improve recognition accuracy.

Hyperparameters of the model will affect the recognition accuracy. We divide the target domain with 40 samples from 1 subject into a verification set and a test set for the specific hyperparameter configuration in the algorithm, which follows similar protocols used in Courty et al. (2016). Among them, the training set is an optional 10 samples, and the test set is the remaining 30 samples. Both validation and test sets have no labels. The validation set data and source domain data are trained together to obtain the best accuracy within the range of hyperparameters, and the range of hyperparameter sets follows (Kerdoncuff et al., 2021). Under the best hyperparameters set, the classification accuracy and F1 measure are used to measure the performance of our proposed algorithm on the test set.

Result

Experiment 1

In Experiment 1, JPDA, JPDA (BMP), SSJPDA, and SSJPDA (BMP) algorithms were used to transfer EEG, PPG, GSR, RES, and four-mode fusion data (ALL) of subjects, respectively. **Table 2** shows the average accuracy and F1-measure of 32 subjects in valence and arousal.

Table 2 shows that in the DEAP dataset, the recognition accuracy of multimodal fusion data is less improved than that of single-mode data recognition. Even in the identification of some modes of JPDA and JPDA-BPM, the accuracy of single-mode is higher than that of multi-mode. However, this phenomenon does not appear in the domain adaptation algorithm using substructure. In the classification of valence and arousal by SSJPDA and SSJPDA-BPM algorithms, the recognition accuracy and F1-measure based on multimodal data are generally higher than that of single-mode data. In the recognition of multimodal data, the recognition accuracy of SSJPDA and SSJPDA-BPM in valence is 14.1 and 19.3% higher than that of JPDA and JPDA-BPM, respectively. In the recognition accuracy of arousal, SSJPDA and SSJPDA-BPM are higher than JPDA and JPDA-BPM by 11.8 and 12.4%, respectively. In the single-mode recognition, SSJPDA-BMP has higher recognition accuracy and F1 than JPDA-BMP in every single mode. Similar rules also appear in the comparison between SSJPDA and JPDA. By comparing the recognition ability of the two transfer learning algorithms with or without the BPM algorithm in each mode, we find that the BPM algorithm is more effective in the transfer learning algorithm with substructure. Among the algorithms that do not use substructures, whether to use the BPM algorithm has little impact on transfer performance.

In order to present the representations generated by different methods more intuitively, we use the t-SNE algorithm in multimodal data experiments to reduce the dimension and visualize the representations generated by different algorithms. **Figure 1** is the t-SNE diagrams of each algorithm in Experiment 1 on multimodal data. The dots legend represents the source domain data, and the legend of the star represents the target domain data. The light blue and dark blue represent positive samples, and the orange and red represent negative samples.

According to **Figure 1**, the representations generated by different algorithms have consistent performance, regardless of valence or arousal classification. The substructures generated by SSJPDA and SSJPDA-BPM through clustering in the domain can significantly reduce the quantity of data. JPDA-BPM and SSJPDA-BPM can lessen the intra-class sample distance and increase the inter-class sample distance in the same domain. At the same time, they can make the same kind of samples in different domains align better compared with not using the BPM algorithm. The representation generated by SSJPDA-BPM has better separability than others.

TABLE 2 The average accuracy (ACC_100%) and F1-measure in different algorithms with single-mode and multi-mode data in valence and arousal classification.

Method	**Modality**	**Valence**		**Arousal**	
		ACC	**F1-measure**	**ACC**	**F1-measure**
JPDA	EEG	0.529	0.563	0.549	0.615
	PPG	0.561	0.603	0.551	0.589
	GSR	0.537	0.578	0.567	0.619
	RES	0.531	0.574	0.509	0.567
	ALL	0.541	0.576	0.568	0.626
JPDA-BPM	EEG	0.536	0.605	0.525	0.624
	PPG	0.536	0.63	0.551	0.582
	GSR	0.553	0.446	0.537	0.57
	RES	0.517	0.555	0.537	0.613
	ALL	0.533	0.615	0.573	0.613
SSJPDA	EEG	0.604	0.617	0.614	0.645
	PPG	0.588	0.537	0.633	0.634
	GSR	0.605	0.596	0.613	0.618
	RES	0.614	0.643	0.619	0.614
	ALL	0.617	0.627	0.635	0.643
SSJPDA-BPM	EEG	0.621	0.645	0.629	0.655
	PPG	0.62	0.619	**0.648**	0.652
	GSR	0.608	0.581	0.62	0.65
	RES	0.595	0.601	0.636	0.653
	ALL	**0.636**	**0.653**	0.644	**0.679**

The numbers in bold indicate the highest value of the experimental results.

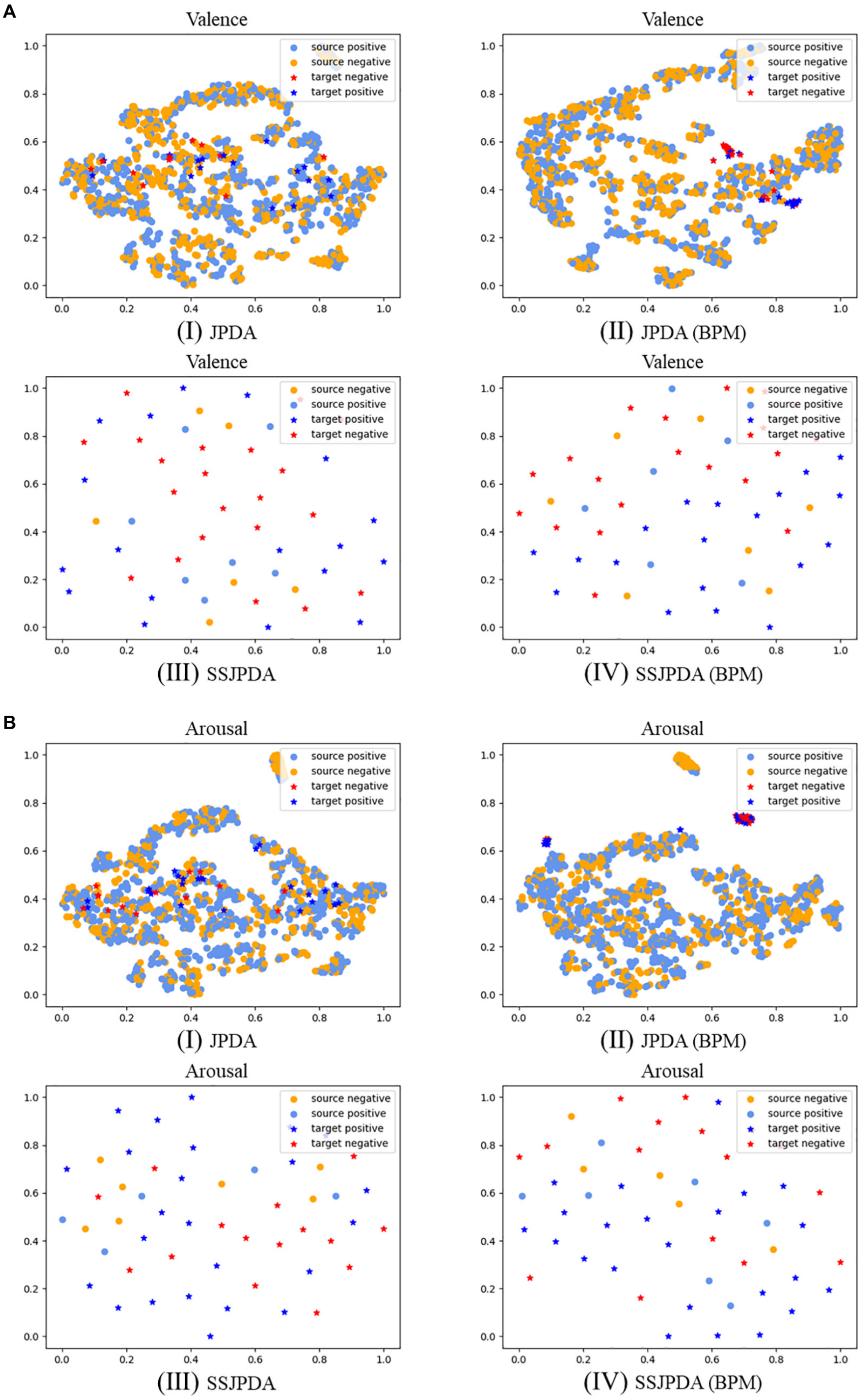


FIGURE 1
The source and target domain's prediction samples are projected to two-dimensional visualization through t-SNE in multimodal data experiments with different algorithms. **(A)** Shows valence classification representations, and **(B)** shows arousal classification representations, where (I) is JPDA algorithm, (II) is JPDA (BPM) algorithm, (III) is SSJPDA algorithm, (IV) is SSJPDA(BPM) algorithm.

Experiment 2

The source domain data and target domain data settings of Experiment 2 are the same as Experiment 1, but only fusion data is used for comparison in the different algorithms. Traditional machine learning and transfer learning algorithms are used to classify valence and arousal. Because the TCA, JDA, BDA, and JPDA algorithms all use the 1-Nearest Neighbor

(1NN) model in classification, we choose 1NN as the traditional machine learning model to compare the impact of the transfer learning algorithm on recognition results. Table 3 shows the average accuracy and F1-measure of 32 subjects using different algorithms in valence and arousal.

Table 3 shows that in the problem of emotion recognition based on the DEAP dataset, when the data distribution of the source domain and target domain is different, the performance of all transfer learning algorithms is better than the 1NN algorithm. In recognition of valence and arousal, the algorithm with the worst classification accuracy in the transfer learning algorithm is still 1.2% (TCA) and 2.2% (JDA) higher than 1NN, respectively. We proposed SSJPDA-BPM algorithm has the best performance. The recognition accuracy and F1-measure values of valence are 63.3 and 65.3%, respectively. The recognition accuracy and F1-measure arousal values are 64.4 and 67.9%, respectively. Its accuracy and F1-measure values are higher than other algorithms. Compared with the traditional transfer learning algorithm, SSJPDA-BPM has higher classification accuracy than TCA, JDA, and BDA by 29.8, 28.2, and 22.5%, respectively, in valence classification. In the recognition accuracy of arousal, SSJPDA-BPM is 23.6, 25.1, and 19.7% higher than TCA, JDA, and BDA, respectively. The comparison results of whether to use BPM and SS algorithms have been described in detail in Experiment 1, which will not be explained in this part.

Figure 2 is the line chart showing the recognition accuracy of each algorithm in Experiment 2 in 32 subjects in descending order, of which Figure 2A is the recognition accuracy of valence and Figure 2B is the recognition accuracy of arousal. The gray horizontal line is the chance level of 50% for the two classes. Each color corresponds to an algorithm. Subjects above the gray level line are represented by upward triangles. The recognition accuracy of this subject in the algorithm is higher than that of the chance level. Downward triangles represent subjects below the gray level line, and the recognition accuracy of this subject in the algorithm is lower than the accuracy of the chance level.

TABLE 3 The average accuracy and F1-measure of different algorithms in valence and arousal classification.

Method	**Valence**		**Arousal**	
	ACC	**F1**	**ACC**	**F1**
1NN	0.484	0.529	0.504	0.555
TCA	0.49	0.533	0.521	0.583
JDA	0.496	0.535	0.515	0.578
BDA	0.519	0.56	0.538	0.572
JPDA	0.541	0.576	0.568	0.626
JPDA-BPM	0.533	0.615	0.573	0.613
SSJPDA	0.617	0.627	0.635	0.643
SSJPDA-BPM	**0.636**	**0.653**	**0.644**	**0.679**

The numbers in bold indicate the highest value of the experimental results.

Figure 2A shows that more than half of the subjects have a recognition accuracy higher than the chance level of 50% for two classes in recognition of valence by the 1NN, TCA, and JDA algorithms. The recognition accuracy of 1NN and TCA in some subjects is less than 30%. Therefore, the average recognition accuracy of these two algorithms is lower than JDA. By comparing JDA, BDA, and JPDA algorithms in order of this arrangement, we can see that the number of people whose three algorithms are higher than the chance level of 50% is slowly increasing. Meanwhile, the highest and lowest recognition accuracy of subjects in the test set is also gradually increasing. The performance of JPDA-BPM is lower than that of JPDA. Although JPDA-BPM algorithm has more subjects with recognition accuracy higher than 70 and 60%, wrong matching still leads to more subjects with recognition accuracy lower than 45%. The SSJPDA and SSJPDA-BPM algorithms have improved compared to the original algorithm. It is worth noting that the recognition accuracy of the SSJPDA-BPM algorithm is above 55% in all subjects.

Figure 2B shows that the number of subjects with arousal recognition accuracy higher than the chance level exceeded half of the total sample size. 1NN, TCA, and JDA algorithms have more than 70% recognition accuracy in some subjects. However, the recognition performance of the algorithm is poor in some subjects, and its recognition accuracy is lower than 35%, which leads to the low average recognition accuracy of these three algorithms. In the JPDA-BPM algorithm, one subject has a recognition accuracy of 85%, which is the highest among the eight algorithms. Meanwhile, its minimum recognition accuracy is 35%, and the number of people lower than the chance level of 50% is also higher than JPDA, which leads to little difference between its average recognition accuracy and JPDA. In the comparison between SSJPDA and SSJPDA-BPM, the performance of SSJPDA-BPM is generally higher than SSJPDA, and the recognition accuracy is lower than SSJPDA only in a few subjects. Comparing whether to use the SS method, SSJPDA and SSJPDA-BPM have been improved compared to the original algorithm, and the recognition accuracy of all subjects is above 50%.

Discussion

The performance of different algorithms with multi modal and single modal data

In the algorithm based on non-substructure, the recognition accuracy using multimodal data is less improved than that using single-mode data. In the substructure-based algorithm, multimodal data can significantly improve recognition performance. Multimodal data in JPDA, the amount of data in the source domain and target domain are very different. The

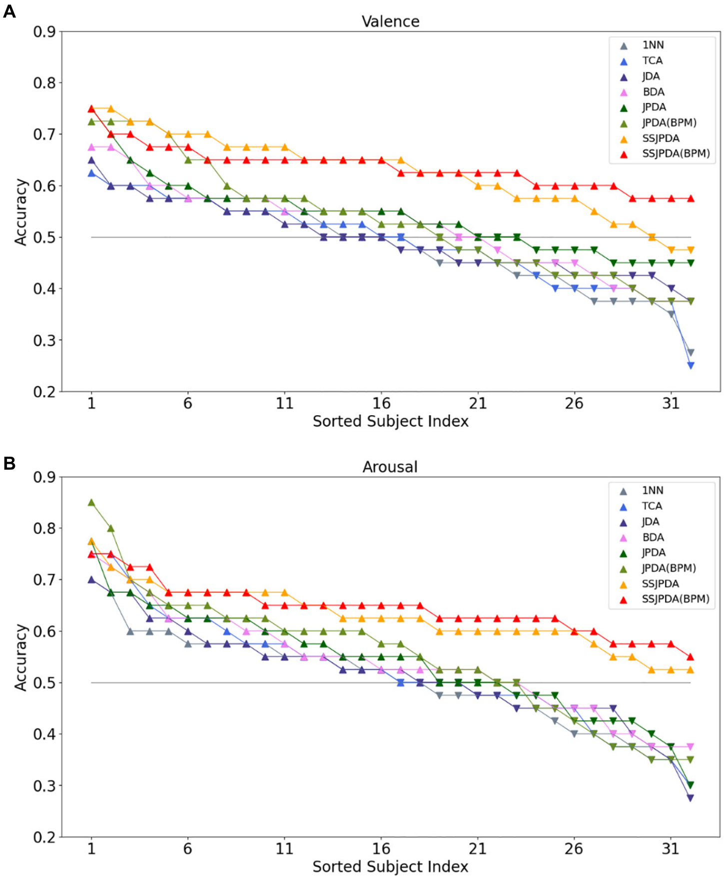


FIGURE 2
The recognition accuracy of each algorithm in Experiment 2 in 32 subjects was ranked in descending order. **(A)** Shows the recognition accuracy of valence in different algorithms of 32 subjects, and **(B)** shows the recognition accuracy of arousal in different algorithms of 32 subjects.

source domain consists of 31 subjects, each of which contains 40 samples. The target domain is 40 samples from one subject, of which 40 samples are also divided into a validation set composed of 10 random samples and a test set consisting of 30 random samples. Therefore, there is an enormous difference in the data volume between the source domain and the target domain. When the source and the target domain are projected to the same feature space, the probability of false matching will increase, which affects transfer recognition's accuracy.

Fusing the features of the four modes will increase the sample dimensions of the source and the target domain. The probability of sample error matching is greater than that of single-mode identification, so the performance of the non-substructure algorithm in multimodal data identification is poor. The transfer learning algorithm based on substructure can avoid error matching caused by sample dimensions increasing and data volume differences between the source and target domains. SSJPDA first generates substructures by clustering in the domain and then matches the substructures. The generation of substructures can dramatically reduce the data volume gap between the source and target domains. This can significantly reduce the probability of false matching. Therefore, the SSJPDA algorithm performs better than JPDA in single-mode emotion recognition. Without the influence of data volume, multimodal fusion data can provide more dimensional information to align the substructures of the source domain and target domains' substructures. Therefore, using the SSJPDA algorithm to recognize multi-mode emotional data can obtain high recognition accuracy.

The application of BPM in SSJPDA can significantly improve recognition performance. Because the emotional labels of subjects in the DEAP dataset are provided by the subjects themselves, this will affect the consistency of the emotional labels of different subjects. At the same time, because the emotional stimulation of the DEAP dataset depends on multimedia clips, some subjects also have the problem of weak emotional stimulation. In this experiment, the source domain contains all the test samples of 31 subjects, so there must be many abnormal samples and noise in the source domain. If no substructure is generated in the source domain and the data is projected directly, the abnormal samples and noise greatly impact the projecting matrix. Therefore, the advantages of the BPM algorithm are not reflected in JPDA. However, the substructure algorithm can cluster the noise or outliers of samples into the substructures of adjacent samples to reduce the impact of noise and outliers. When the source and target domain samples are clustered into substructures, we fully consider the distribution differences between the source and target domain substructures. Projecting the substructure through two different projecting matrices can better project the substructure of the source domain and the target domain to the feature space to improve the algorithm's recognition performance.

The comparison of different algorithms

When the data distribution of the training set and test set is inconsistent, the traditional machine learning algorithm cannot be competent for classification. Therefore, the 1-NearestNeighbor (1NN) algorithm performs worst in this emotion recognition problem. The purpose of transfer learning is to solve the inconsistency between the data distribution of the training set and test set, that is, the inconsistency between the distribution of the source domain and target domain. Therefore, the transfer learning algorithm performs well in this emotion recognition problem. Among them, transfer component analysis (TCA) assumes that if the marginal distributions of the source domain and the target domain are close, the conditional distributions of the two domains will also be close. Therefore, TCA projects the source and target domain data together into a high-dimensional reproducing kernel Hilbert space. In this space, the data distance between the source and the target is minimized, while their respective internal attributes are preserved to the greatest extent to complete the transfer learning. The joint distribution adaptation (JDA) method simultaneously assumes that the marginal and conditional distribution of the source and target domains are different. Then the two distributions are adapted together to achieve transfer. The goal of JDA is to reduce the distance between the source and target domain's joint probability distribution to complete the transfer learning. Balanced distribution adaptation (BDA) is improved on the basis of JDA. BDA assumes that marginal distribution adaptation and conditional distribution adaptation are not equally important. BDA adaptively adjusts the importance of marginal and conditional distribution in the distribution adaptation process according to specific data fields to complete the transfer.

We proposed the SSJPDA algorithm can better measure the distribution difference between the two domains through the joint probability distribution. This is better than JDA and BDA algorithms, which directly calculate the sum of marginal probability and conditional probability distribution differences between the two domains. In the SSJPDA algorithm, the algorithm's transferability is achieved by minimizing the difference in joint probability distribution between different domains of the same class, and the algorithm's discriminability is achieved by maximizing the difference in joint probability distribution between different domains. At the same time, using substructures reduces the difference in data volume between the source domain and the target domain and reduces the impact of noise or outliers. After using the substructure, the SSJPDA-BPM algorithm we proposed fully considers the distribution difference between the substructure of the source domain and the target domain and projects the substructure through two different mapping matrices to improve the performance of the algorithm further. Therefore, this paper's SSJPDA (BMP) algorithm has the highest recognition performance accuracy.

Discussion on negative transfer

Negative transfer means that the knowledge learned in the source domain has a negative effect on the learning in the target domain. When the source domain data is not similar to the target domain data, or the source domain data is similar to the target domain data, but the transfer learning method is not good enough that no transferable components are found, the negative transfer is likely to occur in those two cases (Pan and Yang, 2009). In this experiment, the distribution of source domain data and target domain data are different. Through the multi-source domain transfer method, the data in the target domain is correctly classified by using the knowledge learned from multiple source domains so that the target domain can learn more comprehensive feature information. This can well avoid the negative transfer caused by the low correlation between the source domain and the target domain in the single source domain transfer.

However, if the source domain data used in the transfer learning algorithm contains a lot of noise, it is likely to negatively impact the classification model. The multiple source domain transfer method will further amplify the impact of noise. Regrettably, the four physiological signals, especially EEG signals, in this experiment contain numerous noise and abnormal samples. Therefore, the noise and abnormal samples in the source and target domains will inevitably lead to negative transfer. Therefore, in addition to SSJPDA-BPM algorithm, the classification accuracy of every algorithm in some subjects is lower than the chance level of 50% for two classes.

Compared with other algorithms, SSJPDA and SSJPDA-BPM generate substructures in the source domain and target domain. These substructures can properly process the data according to the data's similarity, which can validly reduce the negative impact of noise and abnormal samples in the source and target domains. It can effectively avoid negative transfer and improve the performance of the transfer learning algorithm. At the same time, as traditional migration learning methods, TCA, JDA, and BDA algorithms have a better effect on the transfer of feature size within a certain threshold. The information redundancy caused by too large feature vectors makes the impact of confusing information greater than that of task-related information, resulting in negative transfer (Zhang et al., 2020). However, SSJPDA and SSJPDA-BPM can filter abnormal samples affected by confusing information through substructure, which further improves the algorithm's performance.

More than that, how to transfer the components found in the source and target domain data also affects the negative transfer. In comparing whether to use the BPM algorithm, if the algorithm finds the correct transferable components, projecting the effective data to the feature space through two different projecting matrices can improve the algorithm's performance and better avoid the negative transfer. However, suppose there is a lot of noise and outliers in the data. In that case, the BPM algorithm changes from an excellent method that avoids more negative transfers to a lousy method that leads to more negative transfers.

Conclusion

This paper proposes SSJPDA and SSJPDA-BPM algorithms to use the labeled physiological data to recognize the emotion of new subjects. We also explored single-mode and multimodal data's influence on emotion recognition based on physiological signals. The performance of the SSJPDA-BPM algorithm is verified by the comparative experiments of various algorithms on DEAP dataset. The results show that SSJPDA and SSJPDA-BPM algorithms can better deal with noise and outliers in data by clustering substructures. Meanwhile, these algorithms can reduce the quantity of data that better use the multi-dimensional information provided by multimodal fusion data. BPM algorithm can project the substructure through two different projecting matrices, which can better project the source domain and target domain data to the feature space, to improve the algorithm's recognition performance. The experimental results show that the average recognition accuracy of the proposed SSJPDA-BPM algorithm in the multimodal fusion physiological data is 63.6 and 64.4% in valence and arousal, respectively.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: http://www.eecs.qmul.ac.uk/mmv/datasets/deap/.

Author contributions

ZF, XH, and JH: conceptualization and supervision. BZ, YL, and HW: data curation. ZF, XH, and BZ: methodology, writing – original draft, and review and editing. BZ, HW, and ZF: validation. XH and YL: visualization. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Technology Innovation Project of Hubei Province of China (Grant No. 2019AEA171) and the Joint fund of Science and Technology Department of Liaoning Province and State Key Laboratory of Robotics, China (Grant No. 2020-KF-22-01).

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

References

Abdulsalam, W. H., Alhamdani, R. S., and Abdullah, M. N. (2019). Facial emotion recognition from videos using deep convolutional neural networks. *Int. J. Mach. Learn. Comput.* 9, 14–19. doi: 10.18178/ijmlc.2019.9.1.759

Al-Nafjan, A., Hosny, M., Al-Ohali, Y., and Al-Wabil, A. (2017). Review and classification of emotion recognition based on EEG brain-computer interface system research: A systematic review. *Appl. Sci.* 7:1239. doi: 10.3390/app7121239

Chai, X., Wang, Q., Zhao, Y., Li, Y., Liu, D., Liu, X., et al. (2017). A fast, efficient domain adaptation technique for cross-domain electroencephalography (EEG)-based emotion recognition. *Sensors* 17:1014. doi: 10.3390/s17051014

Courty, N., Flamary, R., Habrard, A., and Rakotomamonjy, A. (2017). Joint distribution optimal transportation for domain adaptation. *ArXiv* [preprint]. doi: 10.14288/1.0357417

Courty, N., Flamary, R., Tuia, D., and Rakotomamonjy, A. (2016). Optimal transport for domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 1853–1865. doi: 10.1109/TPAMI.2016.2615921

Das, D., and Lee, C. G. (2018). "Unsupervised domain adaptation using regularized hyper-graph matching," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, (Cham: IEEE), 3758–3762. doi: 10.1109/ICIP.2018.8451152

Du, G., Wang, Z., Gao, B., Mumtaz, S., Abualnaja, K. M., and Du, C. (2020). A convolution bidirectional long short-term memory neural network for driver emotion recognition. *IEEE Trans. Intell. Transp. Syst.* 22, 4570–4578. doi: 10.1109/TITS.2020.3007357

Egger, M., Ley, M., and Hanke, S. (2019). Emotion recognition from physiological signal analysis: A review. *Electron. Notes Theor. Comput. Sci.* 343, 35–55. doi: 10.1016/j.entcs.2019.04.009

Fu, Z., He, X., Wang, E., Huo, J., Huang, J., and Wu, D. (2021). Personalized human activity recognition based on integrated wearable sensor and transfer learning. *Sensors* 21:885. doi: 10.3390/s21030885

Hao, Y., Shi, Z., and Liu, Y. (2020). "A Wireless-Vision Dataset for Privacy Preserving Human Activity Recognition," in *2020 Fourth International Conference on Multimedia Computing, Networking and Applications (MCNA)*, (Valencia: IEEE), 97–105. doi: 10.1109/MCNA50957.2020.9264288

He, C., Yao, Y. J., and Ye, X. S. (2017). "An emotion recognition system based on physiological signals obtained by wearable sensors," in *Wearable Sensors and Robots*, eds C. Yang, G. Virk, H. Yang. (Singapore: Springer), 15–25. doi: 10.1007/978-981-10-2404-7_2

He, X., Huang, J., and Zeng, Z. (2021). "Logistic Regression Based Multi-task, Multi-kernel Learning for Emotion Recognition," in *2021 6th IEEE International Conference on Advanced Robotics and Mechatronics (ICARM)*, (Cham: IEEE), 572–577. doi: 10.1109/ICARM52023.2021.9536130

Hosseinifard, B., Moradi, M. H., and Rostami, R. (2013). Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal. *Comput. Methods Programs Biomed.* 109, 339–345. doi: 10.1016/j.cmpb.2012.10.008

Huang, H., Xie, Q., Pan, J., He, Y., Wen, Z., Yu, R., et al. (2019). An EEG-based brain computer interface for emotion recognition and its application in patients with disorder of consciousness. *IEEE Trans. Affect. Comput.* 12, 832–842. doi: 10.1109/TAFFC.2019.2901456

Kerdoncuff, T., Emonet, R., and Sebban, M. (2021). "Metric learning in optimal transport for domain adaptation," in *Proceedings of the twenty-ninth international conference on international joint conferences on artificial intelligence*, Yokohama, 2162–2168. doi: 10.24963/ijcai.2020/295

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

Lawrence, K., Campbell, R., and Skuse, D. (2015). Age, gender, and puberty influence the development of facial emotion recognition. *Front. Psychol.* 6:761. doi: 10.3389/fpsyg.2015.00761

Li, J., Qiu, S., Shen, Y. Y., Liu, C. L., and He, H. (2019). Multisource transfer learning for cross-subject EEG emotion recognition. *IEEE Trans. Cybernetics* 50, 3281–3293. doi: 10.1109/TCYB.2019.2904052

Li, W., Zhang, Z., and Song, A. (2021). Physiological-signal-based emotion recognition: An odyssey from methodology to philosophy. *Measurement* 172:108747. doi: 10.1016/j.measurement.2020.108747

Li, Y., Zhao, T., and Kawahara, T. (2019). "Improved End-to-End Speech Emotion Recognition Using Self Attention Mechanism and Multitask Learning," in *INTERSPEECH*, (Kolkata: ISCA), 2803–2807. doi: 10.21437/Interspeech.2019-2594

Liao, J., Zhong, Q., Zhu, Y., and Cai, D. (2020). Multimodal physiological signal emotion recognition based on convolutional recurrent neural network. *IOP Conf. Series* 782:032005. doi: 10.1088/1757-899X/782/3/032005

Lin, Y. P., and Jung, T. P. (2017). Improving EEG-based emotion classification using conditional transfer learning. *Front. Hum. Neurosci.* 11:334. doi: 10.3389/fnhum.2017.00334

Liu, W., Qian, J., Yao, Z., Jiao, X., and Pan, J. (2019). Convolutional two-stream network using multi-facial feature fusion for driver fatigue detection. *Futur. Internet* 11:115. doi: 10.3390/fi11050115

Liu, W., Zheng, W. L., and Lu, B. L. (2016). "Emotion recognition using multimodal deep learning," in *International Conference on Neural Information Processing*, (Cham: Springer), 521–529. doi: 10.1007/978-3-319-46672-9_58

Long, M., Zhu, H., Wang, J., and Jordan, M. I. (2017). "Deep transfer learning with joint adaptation networks," in *Proceedings of the 34th international conference on machine learning*, Sydney, NSW, 2208–2217. doi: 10.48550/arXiv.1605.06636

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

Lu, W., Chen, Y., Wang, J., and Qin, X. (2021). Cross-domain activity recognition via substructural optimal transport. *Neurocomputing* 454, 65–75. doi: 10.1016/j.neucom.2021.04.124

Mühl, C., Allison, B., Nijholt, A., and Chanel, G. (2014). A survey of affective brain computer interfaces: Principles, state-of-the-art, and challenges. *Brain Comput. Interfaces* 1, 66–84. doi: 10.1080/2326263X.2014.912881

Pan, S. J., Tsang, I. W., Kwok, J. T., and Yang, Q. (2010). Domain adaptation via transfer component analysis. *IEEE Trans. Neural. Netw.* 22, 199–210. doi: 10.1109/TNN.2010.2091281

Pan, S. J., and Yang, Q. (2009). A survey on transfer learning. *IEEE Trans. knowledge Eng.* 22, 1345–1359. doi: 10.1109/TKDE.2009.191

Picard, R. W. (2003). Affective computing: Challenges. *Int. J. Hum. Comput. Stud.* 59, 55–64. doi: 10.1016/S1071-5819(03)00052-1

Piho, L., and Tjahjadi, T. (2018). A mutual information based adaptive windowing of informative EEG for emotion recognition. *IEEE Trans. Affect. Comput.* 11, 722–735. doi: 10.1109/TAFFC.2018.2840973

Rao, V. P., Puwakpitiyage, C. H., Azizi, M. M., Tee, W. J., Murugesan, R. K., and Hamzah, M. D. (2018). "Emotion recognition in e-commerce activities using EEG-based brain computer interface," in *2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA)*, (Manhattan, NY: IEEE), 1–5. doi: 10.1109/ICACCAF.2018.8776818

Shaqra, F. A., Duwairi, R., and Al-Ayyoub, M. (2019). Recognizing emotion from speech based on age and gender using hierarchical models. *Procedia Comput. Sci.* 151, 37–44. doi: 10.1016/j.procs.2019.04.009

Sharot, T., Delgado, M. R., and Phelps, E. A. (2004). How emotion enhances the feeling of remembering. *Nat. Neurosci.* 7, 1376–1380. doi: 10.1038/nn135

Soleymani, M., Villaro-Dixon, F., Pun, T., and Chanel, G. (2017). Toolbox for Emotional feAture extraction from Physiological signals (TEAP). *Front. ICT* 4:1. doi: 10.3389/fict.2017.00001

Song, T., Zheng, W., Lu, C., Zong, Y., Zhang, X., and Cui, Z. (2019). MPED: A multi-modal physiological emotion database for discrete emotion recognition. *IEEE Access* 7, 12177–12191. doi: 10.1109/ACCESS.2019.2891579

Sorkhabi, M. M. (2014). Emotion detection from EEG signals with continuous wavelet analyzing. *Am. J. Comput. Res. Rep.* 2, 66–70. doi: 10.12691/ajcrr-2-4-3

Subha, D. P., Joseph, P. K., Acharya, U. R., and Lim, C. M. (2010). EEG signal analysis: A survey. *J. Meadical Syst.* 34, 195–212. doi: 10.1007/s10916-008-9231-z

Sun, B., Feng, J., and Saenko, K. (2017). "Correlation alignment for unsupervised domain adaptation," in *Domain Adaptation in Computer Vision Applications*, ed. G. Csurka (Cham: Springer), 153–171. doi: 10.1007/978-3-319-58347-1_8

Tian, L., Tang, Y., Hu, L., Ren, Z., and Zhang, W. (2020). Domain adaptation by class centroid matching and local manifold self-learning. *IEEE Trans. Image Proc.* 29, 9703–9718. doi: 10.1109/TIP.2020.3031220

Torres-Valencia, C., Álvarez-López, M., and Orozco-Gutiérrez, Á (2017). SVM-based feature selection methods for emotion recognition from multimodal data. *J. Multimodal User Interfaces* 11, 9–23. doi: 10.1007/s12193-016-0222-y

Verma, G. K., and Tiwary, U. S. (2014). Multimodal fusion framework: A multiresolution approach for emotion classification and recognition from physiological signals. *NeuroImage* 102, 162–172. doi: 10.1016/j.neuroimage.2013.11.007

Wan, Z., Yang, R., Huang, M., Zeng, N., and Liu, X. (2021). A review on transfer learning in EEG signal analysis. *Neurocomputing* 421, 1–14. doi: 10.1016/j.neucom.2020.09.017

Wang, J., Chen, Y., Hao, S., Feng, W., and Shen, Z. (2017). "Balanced distribution adaptation for transfer learning," in *2017 IEEE International Conference on Data Mining (ICDM)*, (Cham: IEEE), 1129–1134. doi: 10.1109/ICDM.2017.150

Wang, J., Chen, Y., Hu, L., Peng, X., and Philip, S. Y. (2018). "Stratified transfer learning for cross-domain activity recognition," in *2018 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, (Cham: IEEE), 1–10. doi: 10.1109/PERCOM.2018.8444572

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

Yin, Z., Wang, Y., Liu, L., Zhang, W., and Zhang, J. (2017). Cross-Subject EEG Feature Selection for Emotion Recognition Using Transfer Recursive

Feature Elimination. *Front. Neurorobot.* 11:19. doi: 10.3389/fnbot.2017.00019

Zhang, W., Deng, L., Zhang, L., and Wu, D. (2020). A survey on negative transfer. *arXiv* [Preprint]. doi: 10.48550/arXiv.2009.00909

Zhang, W., and Wu, D. (2020). "Discriminative joint probability maximum mean discrepancy (DJP-MMD) for domain adaptation," in *2020 International Joint Conference on Neural Networks (IJCNN)*, (Cham: IEEE), 1–8. doi: 10.1109/IJCNN48605.2020.9207365

Zhang, X., Liu, J., Shen, J., Li, S., Hou, K., Hu, B., et al. (2021). Emotion recognition from multimodal physiological signals using a regularized deep fusion of kernel machine. *IEEE Trans. Cybern.* 51, 4386–4399.

Zheng, W. L., and Lu, B. L. (2015). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Trans. Auton. Mental Dev.* 7, 162–175. doi: 10.1109/TAMD.2015.2431497

Zheng, W. L., and Lu, B. L. (2016). "Personalizing EEG-based affective models with transfer learning," in *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, (New York, NY: AAAI Press), 2732–2738.

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 17 October 2022
DOI 10.3389/fnins.2022.1018916

OPEN ACCESS

EDITED BY
Jian Huang,
Huazhong University of Science and Technology, China

REVIEWED BY
Yao Wang,
Tianjin Polytechnic University, China
Yun Lu,
Huizhou University, China

*CORRESPONDENCE
Shixiong Chen
sx.chen@siat.ac.cn
Yuan Tao
taop19772000@163.com

†These authors have contributed equally to this work

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

RECEIVED 14 August 2022
ACCEPTED 26 September 2022
PUBLISHED 17 October 2022

CITATION
Wang X, Zhu M, He Y, Liu Z, Huang X, Pan H, Wang M, Chen S, Tao Y and Li G (2022) Usefulness of phase gradients of otoacoustic emissions in auditory health screening: An exploration with swept tones.
Front. Neurosci. 16:1018916.
doi: 10.3389/fnins.2022.1018916

Usefulness of phase gradients of otoacoustic emissions in auditory health screening: An exploration with swept tones

Xin Wang[1,2,3†], Mingxing Zhu[4†], Yuchao He[1,3], Zhenzhen Liu[5], Xin Huang[6], Hongguang Pan[7], Mingjiang Wang[4], Shixiong Chen[1,3*], Yuan Tao[6*] and Guanglin Li[1,3]

[1]CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, [2]Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen, China, [3]Guangdong-Hong Kong-Macao Joint Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, [4]School of Electronics and Information Engineering, Harbin Institute of Technology, Shenzhen, China, [5]Surgery Division, Epilepsy Center, Shenzhen Children's Hospital, Shenzhen, China, [6]Department of Otorhinolaryngology, Peking University Shenzhen Hospital, Shenzhen, China, [7]Department of Otolaryngology, Shenzhen Children's Hospital, Shenzhen, China

Otoacoustic emissions (OAEs) are low-level sounds generated by the cochlea and widely used as a noninvasive tool to inspect cochlear impairments. However, only the amplitude information of OAE signals is used in current clinical tests, while the OAE phase containing important information about cochlear functions is commonly discarded, due to the insufficient frequency-resolution of existing OAE tests. In this study, swept tones with time-varying frequencies were used to measure stimulus frequency OAEs (SFOAEs) in human subjects, so that high-resolution phase spectra that are not available in existing OAE tests could be obtained and analyzed. The results showed that the phase of swept-tone SFOAEs demonstrated steep gradients as the frequency increased in human subjects with normal hearing. The steep phase gradients were sensitive to auditory functional abnormality caused by cochlear damage and stimulus artifacts introduced by system distortions. At low stimulus levels, the group delays derived from the phase gradients decreased from around 8.5 to 3 ms as the frequency increased from 1 to 10 kHz for subjects with normal hearing, and the pattern of group-delay versus frequency function showed significant difference for subjects with hearing loss. By using the swept-tone technology, the study suggests that the OAE phase gradients could provide highly sensitive information about the cochlear functions and therefore should be integrated into the conventional methods to improve the reliability of auditory health screening.

KEYWORDS

swept-tone, otoacoustic emissions, cochlear impairment, phase gradients, auditory health screening

Introduction

Otoacoustic emissions (OAEs) are low-level sounds generated by the normal activities of the cochlea and can be recorded by a sensitive microphone inside the ear canal. Although it is difficult to physically examine the cochlea due to its deep location inside the temporal bone, the discovery of OAEs provides a non-invasive window to observe the functional status of the cochlea. Studies have shown that OAEs are byproducts of the mechano-electrical activities of outer hair cells (OHCs) that can provide energy feedback to boost the vibrations of the basilar membrane and to amplify the cochlear response to incoming sounds (Dallos et al., 1997). Such physiological OHC activities, also called electromotility, are crucial for the extraordinary frequency selectivity and hearing sensitivity of the human auditory system (Liberman et al., 2002). During the active process of OHC electromotility, part of the extra energy provided by OHCs travels inversely along the basilar membrane and propagates to the ear canal and is recorded as OAEs. A study showed that OAE signals would significantly decrease or even disappear as a result of the blockage of the OHC electromotility (Brownell, 1990). Therefore, the presence of OAEs is a reliable indicator of thriving OHC activities as well as a normal functioning cochlea. Moreover, OAE signals are easy to measure and not affected by attentions or consciousness of the patients (Meric and Collet, 1994). Therefore, OAE measurements have been intensely used in routine hearing screening and audiological assessments in the clinic, especially for the pediatric population that is difficult to test in conventional audiogram assessments (Kemp et al., 1990).

Otoacoustic emission signals are commonly analyzed in the frequency domain by examining both amplitude and phase spectra. Since OHCs at different cochlear positions generate OAEs of different frequencies, their amplitude spectrum has been widely used as a direct approach to inspect the presence of OAE signals and to evaluate OHC functionalities at different frequencies. The OAE amplitude spectrum demonstrates regular spectral periodicity that is unique to each subject (Neumann et al., 1994; Talmadge et al., 1999; Wagner et al., 2008). Another distinctive feature of OAE signals is that the phase changes dramatically with frequency and there is a steep phase gradient in the OAE phase spectrum (Dhar et al., 2002; Choi et al., 2008; Martin et al., 2009; Henin et al., 2011). The steep phase gradient, which reflects the round-trip travel time of OAE signals, is closely related to the active process of OHC electromotility and the frequency selectivity of the cochlea (Shera and Bergevin, 2012). The unique feature of phase gradient enables it to act as an important tool for various studies in the peripheral auditory system (Shera and Guinan, 1999, 2003; Kalluri and Shera, 2001). The phase gradient could also be utilized to investigate the sources and generation mechanisms of different types of OAEs (Zweig and Shera, 1995; Shera and Guinan, 1999; Kalluri and Shera, 2001; Goodman et al., 2003; Lineton and Lutman, 2003), to estimate the cochlear tuning by deducting group delays of the basilar membrane at different frequencies (Shera et al., 2002; Siegel et al., 2005; Shera and Bergevin, 2012), and to examine the olivocochlear efferent control of OHC activities introduced by a contralateral stimulus (Giraud et al., 1995; Guinan et al., 2003). Moreover, a recent study showed that there was a close relation between unstable OAE phase shift and a stiff cochlear partition, suggesting that the phase could be possibly used as a non-invasive way to detect endolymphatic hydrops of Menière's disease (Avan et al., 2011).

Although the OAE phase gradients are useful in different ways, they are mostly restricted to auditory research only and rarely used in clinical practices (Abdala et al., 2018; Liu et al., 2020). Currently, two types of OAEs are measured in the clinics: transient evoked otoacoustic emissions (TEOAEs) measured with brief tones such as clicks, and distortion product otoacoustic emissions (DPOAEs) induced by two sinusoids with closely spaced frequencies. However, only the OAE amplitude is provided upon the completion of both types of measurements while the phase information which essentially represents the OAE signals is completely discarded. One reason is that the measure of phase information requires that the OAEs be tested with a sufficient frequency resolution so that the phase difference between two neighboring frequencies does not exceed 2π to avoid possible phase discontinuities (Shera et al., 2002). However, the frequency resolution of current OAE measurements is usually in the order of hundreds of hertz and it is far from sufficient to capture OAE phases that change dramatically with frequency (Shera and Bergevin, 2012). Recently, Chen et al. (Chen et al., 2013; Jun et al., 2014) proposed a method of using swept tones with rapidly varying frequencies to measure OAE signals with a frequency-resolution as high as a few hertz, making it a great candidate to measure the phase of OAE signals across frequencies. Another possible reason for the lack of use of OAE phases is that the phases of TEOAEs and DPOAEs could be deteriorated by their complex generation mechanisms (Shera and Guinan, 1999). The multiple reflections of TEOAEs (Kemp et al., 1990; Avan et al., 1993; Tognola et al., 1997) and two distinctive sources of DPOAEs (Shera and Guinan, 1999, 2003; Kalluri and Shera, 2001) make it rather difficult to interpret the phase information of the two types of OAEs currently used in the clinic. Stimulus frequency otoacoustic emission (SFOAE) is another type of OAEs commonly evoked by one single stimulus and it attracts increasing attention recently due to its appealing features when compared with other types of OAEs (Guinan et al., 2003; Choi et al., 2008; Bentsen et al., 2011; Cheatham et al., 2011; Ren et al., 2013). Studies showed that SFOAEs were more frequency-specific in reflecting the functional status of corresponding OHCs (Guinan et al., 2003; Choi et al., 2008). It is widely accepted that SFOAEs are generated by linear coherent reflections within the cochlea (Shera and Guinan, 1999;

Goodman et al., 2003; Lineton and Lutman, 2003), and therefore the interpretation of SFOAEs is less complicated, by avoiding multiple reflections in TEOAEs and source mixing in DPOAEs.

The purpose of this study is to use the swept-tone method to measure the phase gradients of SFOAEs in high frequency-resolution, and to demonstrate the role of using phase gradient in combination with conventional amplitude spectrum to obtain more reliable results of auditory health screening in the clinic.

Materials and methods

Subjects

Sixteen subjects were recruited from Shenzhen Institutes of Advanced Technology, with ages from 23 to 36 years old. All subjects declared that they had no congenital auditory disease in the family and no history of ontological surgery or ototoxic drug usage. A conventional audiogram test was also performed on each subject at standard frequencies (1, 2, 3, 4, 6, and 8 kHz) prior to the experiment. Twelve subjects demonstrated normal hearing with thresholds better than 20 dB HL across all frequencies, and the other four subjects had mild hearing loss over certain frequencies due to long-term exposure to loud sounds. The subjects were told to lie comfortably on a foam-covered bed in a double-walled sound booth during the experimental tests. All subjects gave informed consent and provided permission of their data for scientific purposes. The protocol of this study was approved by the Institutional Review Board of Shenzhen Institutes of Advanced Technology (SIAT-IRB-130124-H0015).

Equipment

The presentation of the stimuli and the recording of the acoustic response were controlled by a custom Windows PC program implemented in Matlab environment (Mathworks Inc., USA). Signal Processing Toolbox in Matlab was adopted in this manuscript to further analyze the signals. The digital waveforms of the stimuli were initially synthesized from the PC and then sent to an USB sound card (E-MU 0204, Creative Technology Ltd.) with very low noise background and rather low nonlinear distortion *via* a universal ASIO driver. The sound card is a full duplex with two input channels and four output channels, with all channels of pristine resolutions of 24 bits and very high signal-to-noise ratios (SNRs) up to 117 dB. Two ER-2A earphones (Etymotic Research) were connected to the sound card and converted the digital voltage to acoustic sounds to stimulate the auditory system at the ear canal. The acoustic response was simultaneously recorded by an ER-10+ low-noise microphone (Etymotic Research) seated together with the two earphones inside a foam earplug, and digitized by the USB sound card at a sample rate of 48 kHz. All the original raw data were stored for offline analyses.

Experimental design and procedures

Three-interval protocol to extract stimulus frequency otoacoustic emissions

Since SFOAEs share the same frequency as the evoking stimulus, the extraction strategy becomes more complicated than the simple spectral analysis of TEOAEs or DPOAEs (Kalluri and Shera, 2007). A three-interval protocol (Keefe, 1998; Keefe and Ling, 1998; Chen et al., 2013; Jun et al., 2014) based on the two-tone suppression phenomenon (Kemp and Chum, 1980) was used to extract the SFOAEs in the experiment.

In the three-interval protocol, two stimulus tones (s_1 and s_2) were presented in a specially designed order within three intervals of equal duration (**Figure 1A**). When either stimulus (s_1 or s_2) was presented alone in the first or the second interval, the acoustic response (p_1 or p_2) contained both the stimulus artifact and the evoked SFOAEs. However, when the same stimuli s_1 and s_2 were presented simultaneously during the third interval, the SFOAEs evoked by either stimulus in the response p_{12} would be suppressed by the other stimulus and the amplitude would decrease by $\triangle p_1$ or $\triangle p_2$ as a result. In contrast, the stimulus artifacts related with s_1 and s_2 remained unchanged. When subtracting the response of p_{12} from $p_1 p_2$:

$$\triangle p = p_1 + p_2 - p_{12} \tag{1}$$

Most of the stimulus artifacts would be canceled out, and only the SFOAE amplitude changes in the third interval ($\triangle p_1$ and $\triangle p_2$), as well as other background noises and interferences, would be left in the residue $\triangle p$.

Stimulus generation and presentation

In this study, both stimuli s_1 and s_2 were swept tones with time-varying frequencies to improve the efficiency of SFOAE measurements (**Figure 1B**). The swept tones were constructed by customizing the amplitude and phase spectra in the frequency domain, and converting to the time domain to get the temporal waveform *via* an inverse fast Fourier transform (iFFT) (Müller and Paulo, 2001; Chen et al., 2013).

In the experiment, the duration of all three intervals was kept at 1 s. The frequency of s_1 (the probe tone) was increased linearly from 0.5 to 10 kHz within 1 s, and the frequency of s_2 (the suppressor tone) was kept 200 Hz lower than s_1 (**Figure 1B**). The level of s_1 (L_1) was increased from 45 to 60 dB FPL at a 5-dB step, and the level of remained constant at 80 dB FPL.

Experimental procedures

During the experiment, a foam earplug of the selected size was carefully inserted into the ear canal of the subject. Then

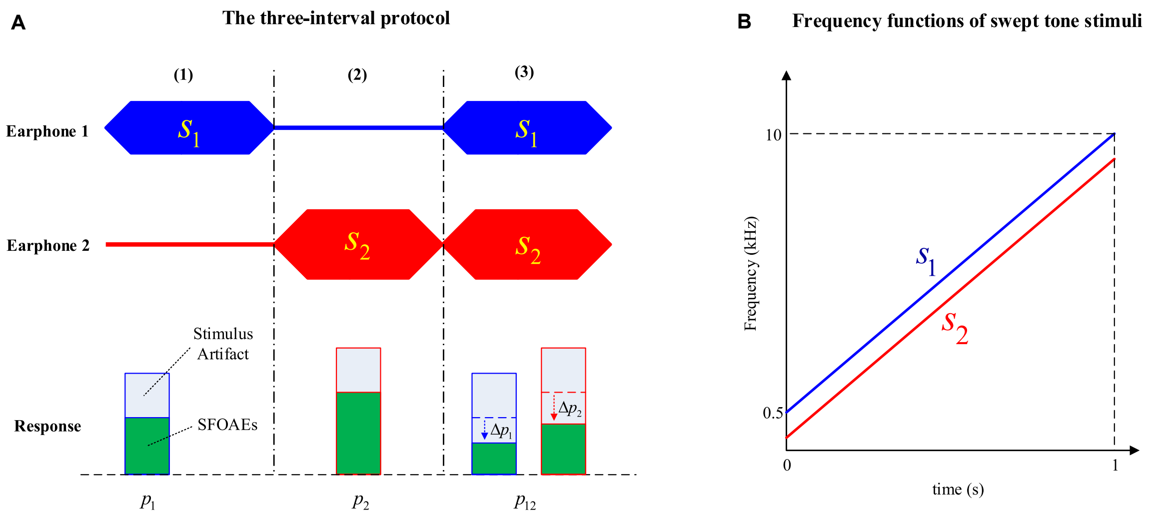


FIGURE 1
The three-interval protocol to measure stimulus frequency otoacoustic emissions (SFOAEs) **(A)** and frequency functions of swept-tone stimuli **(B)**.

the digitally generated swept tones were presented by the two earphones in a three-interval fashion as shown in **Figure 1**, and the acoustic response at the ear canal was simultaneously collected by the microphone seated inside the earplug. For each signal condition, the same stimulus tones were repeatedly presented for 20 times and the responses were digitally averaged to improve the SNR. In the experiments, the stimulus level was equalized across frequencies according to the forward sound pressure (Chen et al., 2014) to avoid the impacts of standing waves. As a comparison, another session with no stimulus level calibration, during which there might be excessive stimulus level within a certain frequency range, was also carried out to examine the effects on the swept-tone SFOAEs.

For verification purposes, the earplug was also inserted into a plastic uniform tube with one end closed and the same procedures were performed to measure the response. The tube was 25 mm in length and 7 mm in diameter, approximately the same size as an adult ear canal. The acoustic response was recorded and analyzed to compare the differences between the human ear and the plastic tube.

Data analysis

A tracking filter that can dynamically follow the instantaneous frequency of the target was used to extract the swept-tone SFOAEs in this study, and the details of the swept-tone and tracking filter could be found in Abdala et al. (2018) and Liu et al. (2020). Another advantage of the tracking filter is that it can easily attenuate the unwanted signal components by placing zeros around the corresponding frequency (Chen et al., 2013). As noted in Eq. 1, two major swept-tone SFOAE components (Δp_1 and Δp_2) coexisted in the residue response Δp. Since nonlinear system distortions might be involved at high stimulus level (80 dB FPL) for s_2, only the swept-tone SFOAEs evoked by s_1 (the Δp_1 component) were analyzed in this study. Accordingly, there was one pole to track the Δp_1 component, and one zero to attenuate the Δp_2 component in the setup of the tracking filter. The signal passed through the tracking filter four times to improve the filter performance (Chen et al., 2013).

As shown in **Figure 2**, the tracking filter was applied to the temporal waveform of Δp to get a dynamic estimate (Δp_1) of the swept-tone SFOAE component Δp_1. Then a fast Fourier transform (FFT), with a fixed 1-s length Hanning window, was performed on the Δp_1, and the magnitude of the FFT result $X_{\Delta p_1{}'}$ was taken as the amplitude spectrum of the swept-tone SFOAEs. Meanwhile, the center frequency of the tracking filter was set 100 Hz above the frequency of Δp_1, and the amplitude of the filter output was calculated as the reference of the noise floor.

To get the phase of the swept-tone SFOAEs, the same tracking filter that was used to obtain $\Delta p_1{}'$ was applied to the response of p_1 in **Figure 1**, and an FFT was performed on the corresponding filter output to get the spectral complex X_{p_1} (**Figure 2**). Then the phase of X_{p_1} was subtracted from the phase of $X_{\Delta p_1{}'}$ (FFT result of $\Delta p_1{}'$), and the unwrapped phase difference was calculated as the phase of the swept-tone SFOAEs. The phase subtraction was used to eliminate the delays introduced by the recording system and the tracking filter, so that the phase gradient attribute to the swept-tone SFOAEs could be truly revealed.

Since the frequency of the stimulus continuously changed with time, the evoked swept-tone SFOAEs were also continuous in frequency, making it possible to obtain the phase spectrum of SFOAEs in high definition to avoid possible discontinuities. For

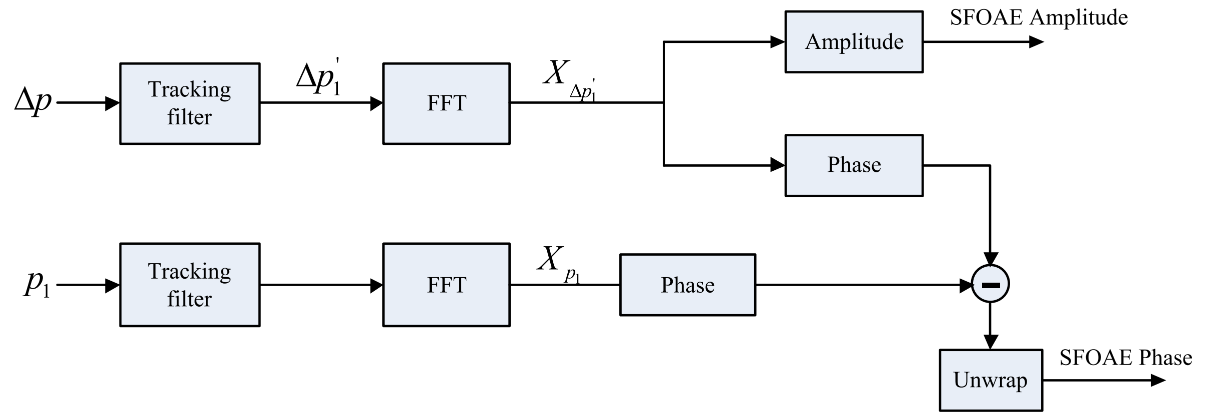


FIGURE 2
Signal processing procedures to get the amplitude and phase of swept-tone stimulus frequency otoacoustic emissions (SFOAEs). The phase subtraction in the last step was to cancel the phase shifts introduced by the stimulus and the tracking filter.

an SFOAE phase spectrum $\phi(f)$ prepared by the procedures in Figure 2, the group delay τ, defined as the transit time of a signal through a system, could be calculated by:

$$\tau = -\frac{1}{2\pi}\frac{d\phi f}{df} \quad (2)$$

The group delay of SFOAEs was a rather useful measure of the travel time of the OAE signals inside the cochlea, and it could provide quite useful information about the functional status of the OHCs and the sharp tuning of the cochlea (Shera and Guinan, 2003; Siegel et al., 2005). In this study, the group delays were calculated at discrete frequencies (f_i) from 1 to 10 kHz (at a 1-kHz step) for all the subjects. For each discrete frequency f_i, the phase-frequency function from $f_i-100Hz$ to $f_i+100Hz$ was fitted with a straight line and the slope was used to calculate the group delay according to Eq. 2.

Results

Swept-tone stimulus frequency otoacoustic emissions in subjects with normal hearing

The presence of OAE signals is a distinctive feature of the healthy human ear that makes it different from other passive systems such as acoustic tubes. The use of swept tones made it possible to observe the OAE features in such a high definition. For comparison purposes, swept-tone SFOAEs were measured in both human ears with normal hearing and a plastic tube of similar sizes under the same signal conditions (L_1 = 50 dB FPL) in this study. A typical comparison of the spectrogram (energy distribution as a function of time and frequency), amplitude and phase spectra between the two responses (represented by Δp in Eq. 1) was shown in Figure 3. The most important finding was that the responses in the human ear and plastic tube showed dramatically different patterns. For the response in the human ear, two ascending lines, which reflected the energy concentrations in the residue response, were clearly observed in the spectrogram (Figure 3A). The two lines corresponded to the two SFOAE components (Δp_1 and Δp_2) and had quite similar frequency patterns as their evoking stimuli (Figure 1B). Then a tracking filter was applied to extract the Δp_1 component, and the filtered amplitude and phase spectra were shown in Figures 3B,C, respectively. The SFOAE amplitude in Figure 3B consisted of slow baseline variations and rapid spectral periodicity (or fine structures) indicated by alternating peaks and troughs. The overall SFOAE amplitude could get around 30 dB FPL above the noise floor for the subject with normal hearing. The pattern of the fine structures is unique for each specific subject. Another unique feature of the ear response was that the phase decreased dramatically as the frequency increased: the amount of the phase decrease could exceed 150 rad when the frequency increased from 0.5 to 10 kHz. However, for the response in the plastic tube, no swept-tone SFOAEs were observed in either the spectrogram (Figure 3D) or the amplitude spectrum (Figure 3E). The phase no longer demonstrated steep gradient and fluctuated around 0 as the frequency increased (Figure 3F).

Swept-tone stimulus frequency otoacoustic emissions in subjects with mild hearing loss

Since SFOAEs are closely related to the normal functions of the cochlea, any alterations in the cochlear functions accompanied by auditory functional abnormality would result in changes in SFOAEs. A typical example of the swept-tone

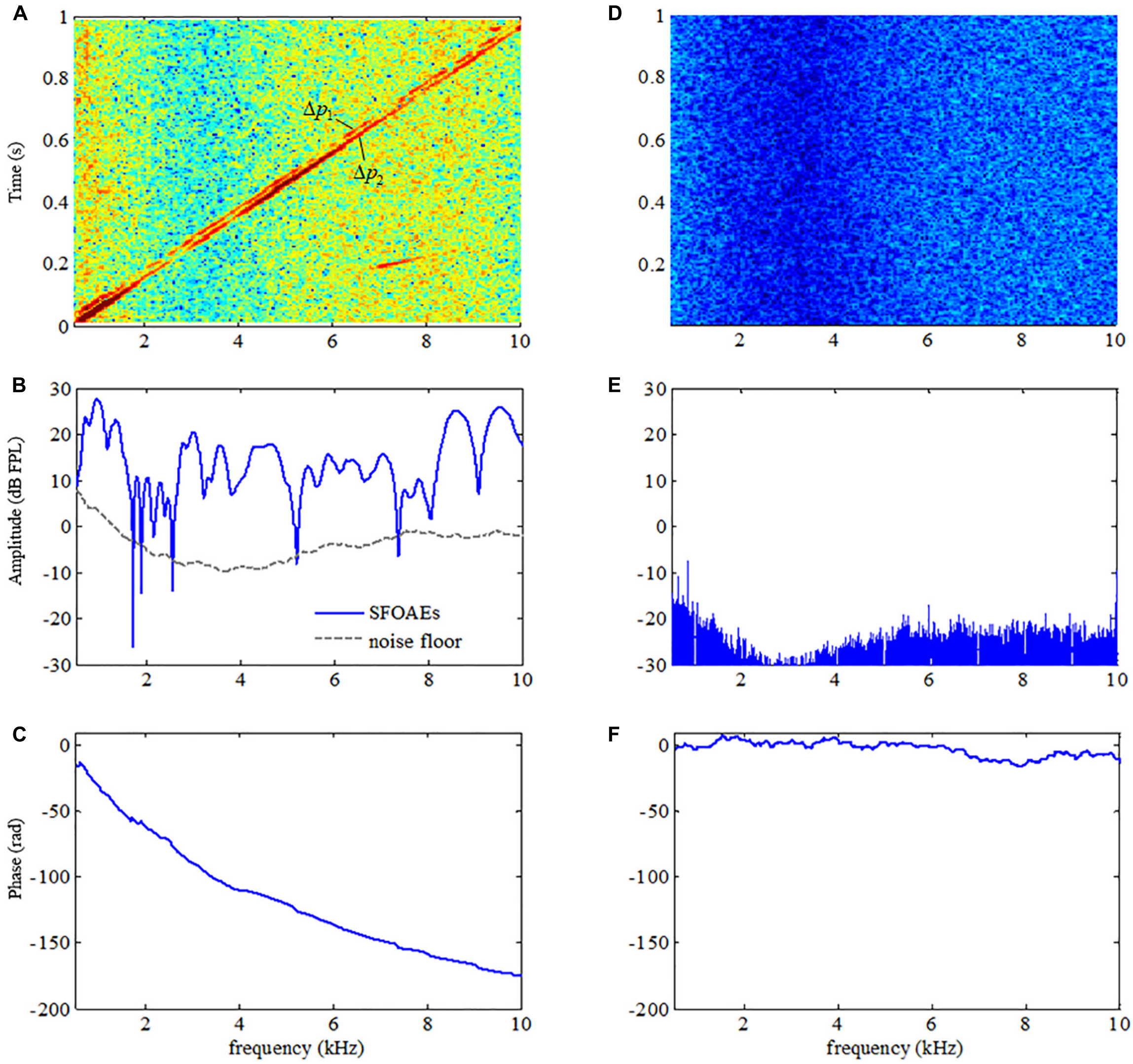


FIGURE 3
Difference in the spectrograms, amplitude, and phase spectra between the acoustic responses in a human ear (**A–C**, respectively) and in a plastic tube (**D–F**, respectively).

SFOAEs of a subject with mild hearing loss of 2–3 kHz was shown in **Figure 4**. As shown in **Figure 4A**, the overall baseline amplitude of the SFOAEs fell below the noise floor within 2–3 kHz, which was consistent with the frequency region of the hearing loss. However, there was a large amplitude of SFOAEs over other frequencies. For the phase spectrum in **Figure 4B**, although steep phase gradients could be observed at most frequencies, the phase function became rather flat when the frequency was from 2 to 3 kHz. The flattening of the phase was consistent with the SFOAE amplitude reduction, as well as the region of the hearing loss. The abrupt phase discontinuities at other frequencies (such as 5–7.3 kHz) were due to the lower SNR at the trough of the SFOAE fine structures where the phase estimation was more susceptible to random noises.

Phase gradients of swept-tone stimulus frequency otoacoustic emissions with artifacts

The SFOAEs of different frequencies originate from activities of OHCs at different positions along the cochlea. In practice, equalization of stimulus level across frequencies was usually desired so that SFOAEs from different frequencies could be compared. However, it is very common to have excessive stimulus level over certain frequency ranges without calibrations of stimulus levels, leading to large artifacts that could not be ignored in the SFOAE analyses (a typical example was shown in **Figure 5**). As shown in **Figure 5B**, a large amplitude of SFOAEs could be observed across all frequencies from 0.5

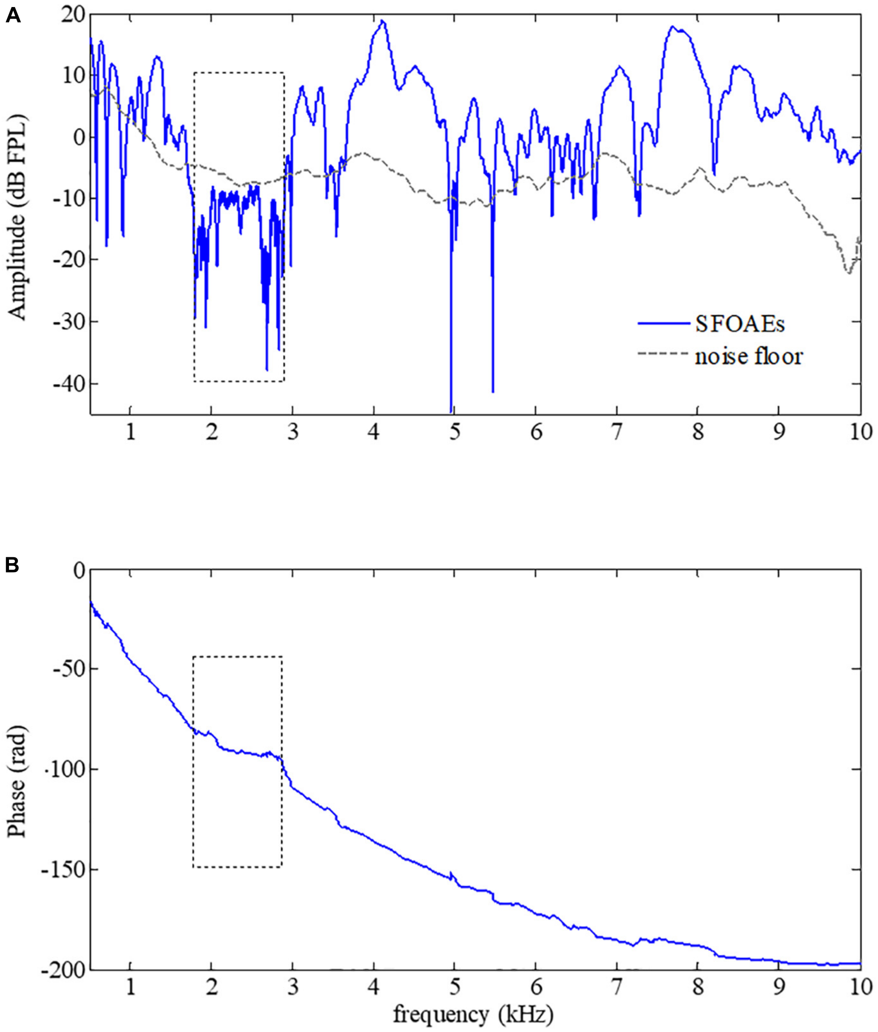


FIGURE 4
Amplitude **(A)** and phase **(B)** of swept-tone stimulus frequency otoacoustic emissions (SFOAEs) of a subject with hearing loss from 2 to 3 kHz (indicated by dotted boxes).

to 10 kHz, and it could reach up to 25 dB FPL around the frequency of 9 kHz. However, as we checked the corresponding stimulus level in **Figure 5A**, it was found that the stimulus level above 5 kHz was much higher than the expected level of 50 dB FPL. Such excessive stimulus level (as high as 20 dB) could introduce nonlinear system distortions that could not be canceled out during the subtraction in Eq. 1 or the filtering by the tracking filter, resulting in unexpected artifacts during the SFOAE measurements (Whitehead et al., 1995; McCreery et al., 2009; Burke et al., 2010; Chen et al., 2014). Fortunately, such artifacts could be correctly identified by the abnormal phase gradient in **Figure 5C**, where the phase gradually violated the steep gradient pattern of OAE signals above 5 kHz. The flattening of the phase curve suggested that there might be a large amplitude of system distortion involved, leading to unreliable results in the extracted signals above 5 kHz.

Groups delays of swept-tone stimulus frequency otoacoustic emissions

Group delays, obtained from the derivative of the phase versus frequency function (Eq. 2), could provide a non-invasive tool to monitor the cochlear tuning that is important for cochlear healthiness (Shera and Bergevin, 2012) as well as the frequency selectivity which is essential for speech perception (Evans, 1975). In this study, the SFOAEs were measured using swept tones to provide SFOAE phases in high frequency-resolution, making it possible to obtain reliable SFOAE group delays with high efficiency. In the experiment, the group delays of the swept-tone SFOAEs were measured at 10 discrete frequencies (from 1 to 10 kHz at a step of 1 kHz) under different stimulus levels. The means and standard deviations of the group delays averaged across the 12 normal-hearing

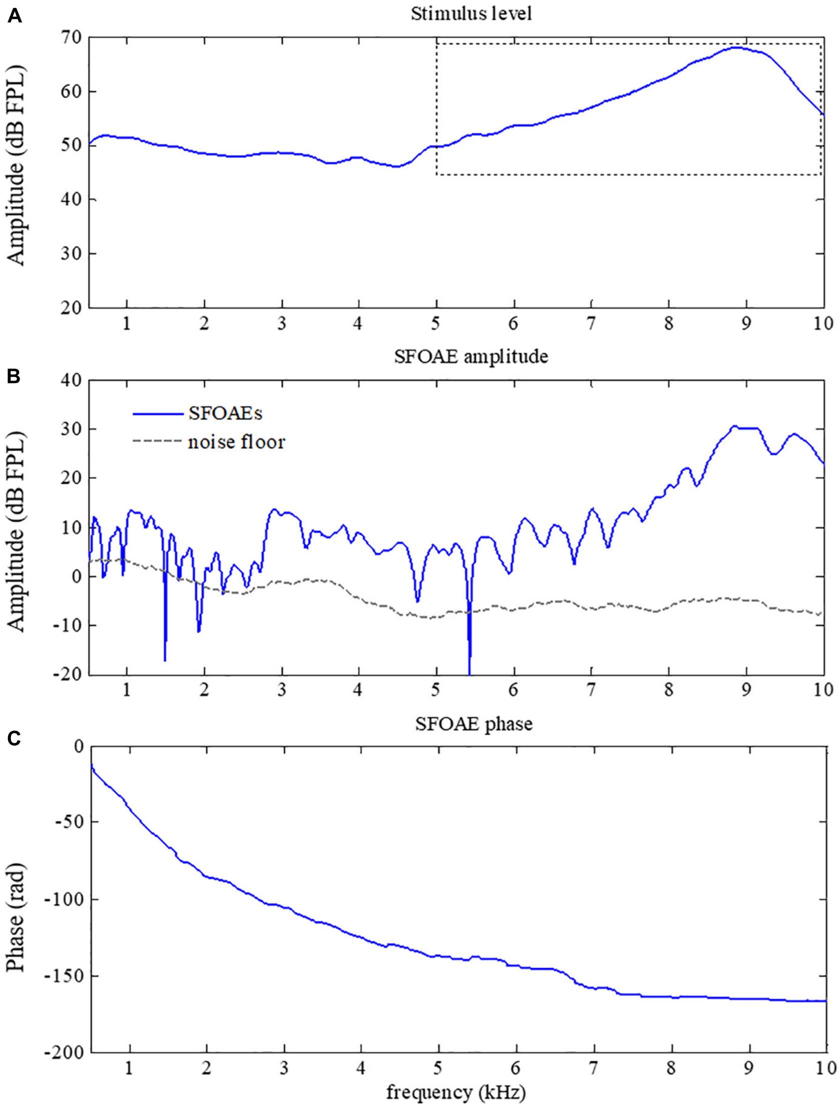


FIGURE 5
Amplitude **(B)** and phase **(C)** of swept-tone stimulus frequency otoacoustic emissions (SFOAEs) with artifacts above 5 kHz caused by excessive stimulus level (indicated by the dotted box in **A**).

subjects were shown in **Figure 6**. For the stimulus of 45 dB FPL, the average group delay decreased from around 8.5 ms to about 3 ms as the frequency increased from 1 to 10 kHz. The group delay decreased more rapidly at lower frequencies. When the stimulus level increased from 45 to 60 dB FPL, the group delay at a given frequency decreased monotonously, which was consistent with the findings of relevant studies (Shera and Guinan, 2003; Bentsen et al., 2011; Shera and Bergevin, 2012). Due to the impacts of low-frequency noises in the ear canal, the group delays of different stimulus levels almost overlapped at 1 kHz. Since the phase versus frequency function became rather flat as consequence of hearing loss (**Figure 4**), the estimated group delays within the corresponding frequency range would approach 0 ms and the group delay pattern would be quite different from **Figure 6** for subjects with hearing loss.

Discussion

Usefulness of stimulus frequency otoacoustic emission phase gradients in auditory health screening

The study showed that the phase gradients of swept-tone SFOAEs were rather useful to help improve the reliability of auditory health screening using OAE measurements. As

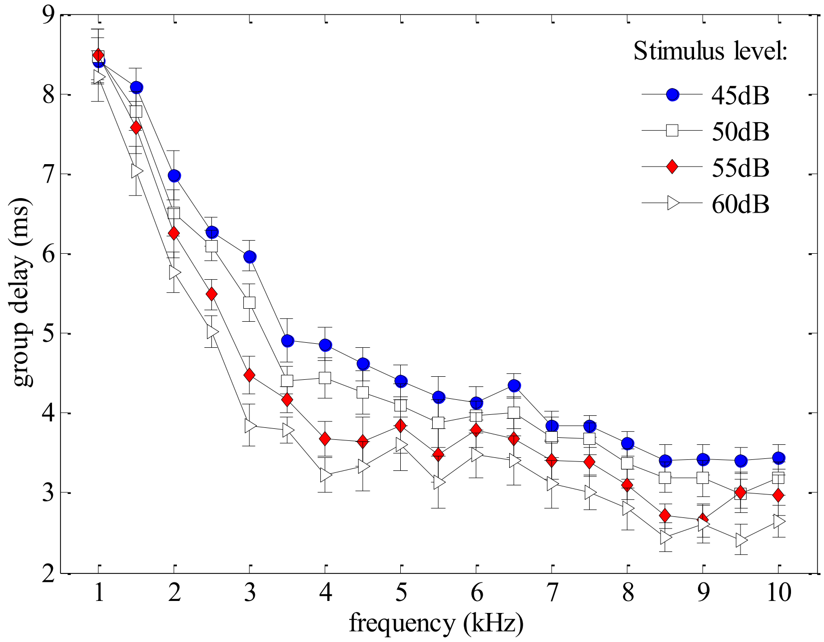


FIGURE 6
Group delays averaged across all subjects as functions of frequency and stimulus level derived from phase gradients of the swept-tone stimulus frequency otoacoustic emissions (SFOAEs).

observed in **Figure 3**, the steep phase gradient of OAE signals is a unique feature of the normal functioning human cochlea. Such steep phase gradient would totally disappear if measurements were made in a passive tube (**Figure 3**) or partially vanish if auditory functional abnormality like hearing loss existed within certain frequency ranges (**Figure 4**). The close relation between the OAE phase gradients and cochlea healthiness is also reported in relevant studies using fixed-frequency tones (Martin et al., 2009; Avan et al., 2011; Shera and Bergevin, 2012). However, the phase information of OAEs is habitually abandoned due to insufficient frequency-resolution in clinical applications. Currently, only the results of OAE amplitude vs. noise level were provided and the presence of OAEs is determined by whether the OAE amplitude is above the noise level (Kemp et al., 1990; Erenberg et al., 1999; Abujamra et al., 2013). One problem is that the screening results are largely dependent on the reliability of the noise estimate. Since there is no universal standard for noise estimation, different levels of noise might be obtained if different algorithms are used for the noise calculation (Attias et al., 2001; Reavis et al., 2008), leading to inconsistent results among different methods. In this study, the noise floor was obtained by setting the center frequency of the tracking filter 100 Hz above the swept-tone SFOAEs (**Figures 3–5**). However, slight differences in the estimated noise floor might be expected if the tracking filter was set in different ways, which might result in differences in the identification of a possible hearing-loss region in **Figure 4A**. In contrast, the calculation of SFOAE phases is independent of the noise estimation, making it more suitable and reliable to indicate the presence of OAE signals or the existence of possible auditory healthy issues (hearing loss in this study) over certain frequencies (**Figure 4B**). However, it is recommended that the phase of SFOAEs should be used in combination with the conventional amplitude spectra for more accurate auditory health screening in clinical practices. Moreover, the group delays calculated from the phase gradient could also be used to detect the abnormality of cochlear tuning that demonstrates evident alterations at the early stage of auditory functional disorders such as hearing loss (Francis and Guinan, 2010; Shera and Bergevin, 2012).

Phase gradients and group delays to detect otoacoustic emission artifacts

The present study also showed that the SFOAE phase gradients could help to identify possible artifacts that otherwise could be falsely treated as actual OAE signals. As mentioned earlier, the major difference between OAE signals and other responses is that the unique phase gradient (**Figure 3**) is so steep that the derived group delay is about 8.5 ms at 1 kHz for all subjects (**Figure 6**), not including the system round-trip delay that the stimulus spending on traveling along the outer and middle ears. The steep phase gradient, as well as the group delay, originates from the signal front delay which is the time difference between the onset of the basilar membrane (BM) and stapes, and the filter delay that the BM spends on building the peak of the traveling wave (Ruggero, 2004). The filter delay of the BM, the major portion of the OAE group delay, is a unique physiological parameter closely related to the frequency selectivity of the cochlear tuning, with sharper tuning corresponding to longer group delays (Shera and Guinan, 2003; Ren et al., 2006). If there is no cochlear tuning involved (such as the response in a passive tube), the steep phase gradient would disappear (**Figure 3**) and the group delay would approach 0 as a consequence. In **Figure 5**, system distortions were involved above 5 kHz due to the excessive stimulus level, leading to incomplete cancellation of stimulus artifacts during the subtraction in Eq. 1. The remaining stimulus artifacts with zero group delay would dominate the low-level SFOAE components ($\triangle p_1$ and $\triangle p_2$) in the residual response $\triangle$p (Eq. 1), resulting in the violation of steep phase gradients at the corresponding frequency range (**Figure 5C**). However, it would be mistaken if we determined the presence of large OAE signals by merely checking the high SNR in the amplitude spectrum above 5 kHz (**Figure 5B**). Therefore, checking the steep phase gradient in combination with the amplitude spectrum is a more reliable way to distinguish OAE signals from other irrelevant interferences or unexpected noises.

Conclusion

In this paper, SFOAEs were measured with swept tones in high frequency-resolution, so that the phase spectrum that is conventionally not feasible in auditory research or clinical tests could be obtained in a quite efficient way. The results demonstrated that the SFOAE phases in human ears showed steep gradients as the frequency increased, and such steep gradients are unique features that make the human-ear response different from other passive systems. The steep phase gradients could help to efficiently validate frequency regions of auditory functional abnormality, and to identify stimulus artifacts that could be mistakenly treated as evident OAE signals in practical applications. The pattern of the group delays derived from SFOAE phase gradients might be used to reflect the cochlear latency characteristics that were useful to evaluate the sharpness of the cochlear tuning and the normalcy of the cochlear frequency selectivity. The study suggested that using swept tones to measure SFOAEs and involving the phase information in combination with the amplitudes could be a rather promising approach to help improve the reliability of current hearing screening in the clinic.

Data availability statement

The datasets presented in this article are available on reasonable request from the corresponding authors. Requests to access the datasets should be directed to SC.

Ethics statement

The protocol of this study was approved by the Institutional Review Board of Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences (SIAT- IRB-130124-H0015). The patients/participants provided their written informed consent to participate in this study.

Author contributions

SC and YT: conceptualization. XW and MZ: data curation. SC, GL, and MZ: funding acquisition. XW and SC: methodology, writing—original draft, and writing—review and editing. SC: supervision. XW, YH, and ZL: validation. XW, XH, and HP: visualization. All authors contributed to the article and approved the submitted version.

Funding

This work was supported in part by the Key-Area Research and Development Program of Guangdong Province (#2020B0909020004), National Natural Science Foundation of China (#81927804 and #62101538), Shenzhen Governmental Basic Research Grant (#JCYJ20180507182241622), Science and Technology Planning Project of Shenzhen (#JSGG20210713091808027 and #JSGG20211029095801002), China Postdoctoral Science Foundation (2022M710968), SIAT Innovation Program for Excellent Young Researchers (E1G027), and CAS President's International Fellowship Initiative Project (2022VEA0012).

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

References

Abdala, C., Ortmann, A. J., and Shera, C. A. (2018). Reflection- and distortion-source otoacoustic emissions: Evidence for increased irregularity in the human cochlea during aging. *J. Assoc. Res. Otolaryngol.* 19, 493–510. doi: 10.1007/s10162-018-0680-x

Abujamra, A. L., Escosteguy, J. R., Dall'Igna, C., Manica, D., Cigana, L. F., Coradini, P., et al. (2013). The use of high-frequency audiometry increases the diagnosis of asymptomatic hearing loss in pediatric patients treated with cisplatin-based chemotherapy. *Pediatr. Blood Cancer* 60, 474–478. doi: 10.1002/pbc.24236

Attias, J., Horovitz, G., El-Hatib, N., and Nageris, B. (2001). Detection and clinical diagnosis of noise-induced hearing loss by otoacoustic emissions. *Noise Health* 3, 19–31.

Avan, P., Bonfils, P., Loth, D., and Wit, H. P. (1993). Temporal patterns of transient-evoked otoacoustic emissions in normal and impaired cochleae. *Hear. Res.* 70, 109–120. doi: 10.1016/0378-5955(93)90055-6

Avan, P., Giraudet, F., Chauveau, B., Gilain, L., and Mom, T. (2011). Unstable distortion-product otoacoustic emission phase in Meniere's disease. *Hear. Res.* 277, 88–95. doi: 10.1016/j.heares.2011.03.006

Bentsen, T., Harte, J. M., and Dau, T. (2011). Human cochlear tuning estimates from stimulus-frequency otoacoustic emissions. *J. Acoust. Soc. Am.* 129, 3797–3807. doi: 10.1121/1.3575596

Brownell, W. E. (1990). Outer hair cell electromotility and otoacoustic emissions. *Ear. Hear.* 11:82. doi: 10.1097/00003446-199004000-00003

Burke, S. R., Rogers, A. R., Neely, S. T., Kopun, J. G., Tan, H., and Gorga, M. P. (2010). Influence of calibration method on distortion-product otoacoustic emission measurements: I. test performance. *Ear. Hear.* 31, 533–545. doi: 10.1097/AUD.0b013e3181d86b3d

Cheatham, M. A., Katz, E. D., Charaziak, K., Dallos, P., and Siegel, J. H. (2011). Using stimulus frequency emissions to characterize cochlear function in mice. *AIP Conf. Proc.* 1403, 383–388. doi: 10.1063/1.3658115

Chen, S., Deng, J., Bian, L., and Li, G. (2013). Stimulus frequency otoacoustic emissions evoked by swept tones. *Hear. Res.* 306, 104–114. doi: 10.1016/j.heares.2013.09.016

Chen, S., Zhang, H., Wang, L., and Li, G. (2014). An in-situ calibration method and the effects on stimulus frequency otoacoustic emissions. *Biomed. Eng. Online* 13:95. doi: 10.1186/1475-925X-13-95

Choi, Y. S., Lee, S. Y., Parham, K., Neely, S. T., and Kim, D. O. (2008). Stimulus-frequency otoacoustic emission: Measurements in humans and simulations with an active cochlear model. *J. Acoust. Soc. Am.* 123, 2651–2669. doi: 10.1121/1.2902184

Dallos, P., He, D. Z. Z., Lin, X., Sziklai, I., Mehta, S., and Evans, B. N. (1997). Acetylcholine, outer hair cell electromotility, and the cochlear amplifier. *J. Neurosci.* 17, 2212–2226. doi: 10.1523/JNEUROSCI.17-06-02212.1997

Dhar, S., Talmadge, C. L., Long, G. R., and Tubis, A. (2002). Multiple internal reflections in the cochlea and their effect on DPOAE fine structure. *J. Acoust. Soc. Am.* 112, 2882–2897. doi: 10.1121/1.1516757

Erenberg, A., Lemons, J., Sia, C., Tunkel, D., Ziring, P., Adams, M., et al. (1999). Newborn and infant hearing loss: Detection and intervention. *Pediatrics* 103, 527–530. doi: 10.1542/peds.103.2.527

Evans, E. (1975). The sharpening of cochlear frequency selectivity in the normal and abnormal cochlea. *Audiology* 14, 419–442. doi: 10.3109/00206097509071754

Francis, N. A., and Guinan, J. J. (2010). Acoustic stimulation of human medial olivocochlear efferents reduces stimulus-frequency and click-evoked otoacoustic emission delays: Implications for cochlear filter bandwidths. *Hear. Res.* 267, 36–45. doi: 10.1016/j.heares.2010.04.009

Giraud, A. L., Collet, L., Chery-Croze, S., Magnan, J., and Chays, A. (1995). Evidence of a medial olivocochlear involvement in contralateral suppression of otoacoustic emissions in humans. *Brain Res.* 705, 15–23. doi: 10.1016/0006-8993(95)01091-2

Goodman, S. S., Withnell, R. H., and Shera, C. A. (2003). The origin of SFOAE microstructure in the guinea pig. *Hear. Res.* 183, 7–17. doi: 10.1016/s0378-5955(03)00193-x

Guinan, J. J. Jr., Backus, B. C., Lilaonitkul, W., and Aharonson, V. (2003). Medial olivocochlear efferent reflex in humans: Otoacoustic emission (OAE) measurement issues and the advantages of stimulus frequency OAEs. *J. Assoc. Res. Otolaryngol.* 4, 521–540. doi: 10.1007/s10162-002-3037-3

Henin, S., Thompson, S., Abdelrazeq, S., and Long, G. R. (2011). Changes in amplitude and phase of distortion-product otoacoustic emission fine-structure and separated components during efferent activation. *J. Acoust. Soc. Am.* 129, 2068–2079. doi: 10.1121/1.3543945

Jun, D., Shixiong, C., Xiaoping, Z., and Guanglin, L. (2014). Using a dynamic tracking filter to extract distortion-product otoacoustic emissions evoked with swept-tone signals. *IEEE J. Biomed. Health Inf.* 18, 1186–1195. doi: 10.1109/JBHI.2013.2285558

Kalluri, R., and Shera, C. A. (2001). Distortion-product source unmixing: A test of the two-mechanism model for DPOAE generation. *J. Acoust. Soc. Am.* 109, 622–637. doi: 10.1121/1.1334597

Kalluri, R., and Shera, C. A. (2007). Comparing stimulus-frequency otoacoustic emissions measured by compression, suppression, and spectral smoothing. *J. Acoust. Soc. Am.* 122, 3562–3575. doi: 10.1121/1.2793604

Keefe, D. H. (1998). Double-evoked otoacoustic emissions. I. Measurement theory and nonlinear coherence. *J. Acoust. Soc. Am.* 103, 3489–3498.

Keefe, D. H., and Ling, R. (1998). Double-evoked otoacoustic emissions. II. Intermittent noise rejection, calibration and ear-canal measurements. *J. Acoust. Soc. Am.* 103, 3499–3508. doi: 10.1121/1.423058

Kemp, D. T., and Chum, R. A. (1980). "Observations on the generator mechanism of stimulus frequency acoustic emissions – Two tone suppression," in *Psychophysical, physiological and behavioural studies in hearing*, eds G. van den Brink and F. A. Bilsen (Delft: Delft Univ. Press). doi: 10.1016/s0378-5955(98)00045-8

Kemp, D. T., Ryan, S., and Bray, P. (1990). A guide to the effective use of otoacoustic emissions. *Ear. Hear.* 11, 93–105. doi: 10.1097/00003446-199004000-00004

Liberman, M. C., Gao, J., He, D. Z. Z., Wu, X., Jia, S., and Zuo, J. (2002). Prestin is required for electromotility of the outer hair cell and for the cochlear amplifier. *Nature* 419, 300–304. doi: 10.1038/nature01059

Lineton, B., and Lutman, M. E. (2003). A parametric model of the spectral periodicity of stimulus frequency otoacoustic emissions. *J. Acoust. Soc. Am.* 114, 883–895. doi: 10.1121/1.1582176

Liu, Y.-W., Kao, S.-L., Wu, H.-T., Liu, T.-C., Fang, T.-Y., and Wang, P.-C. (2020). Transient-evoked otoacoustic emission signals predicting outcomes of acute sensorineural hearing loss in patients with Meniere's disease. *Acta Otolaryngol.* 140, 230–235. doi: 10.1080/00016489.2019.1704865

Martin, G. K., Stagner, B. B., Fahey, P. F., and Lonsbury-Martin, B. L. (2009). Steep and shallow phase gradient distortion product otoacoustic emissions arising basal to the primary tones. *J. Acoust. Soc. Am.* 125, EL85–EL92. doi: 10.1121/1.3073734

McCreery, R. W., Pittman, A., Lewis, J., Neely, S. T., and Stelmachowicz, P. G. (2009). Use of forward pressure level to minimize the influence of acoustic standing waves during probe-microphone hearing-aid verification. *J. Acoust. Soc. Am.* 126, 15–24. doi: 10.1121/1.3143142

Meric, C., and Collet, L. (1994). Attention and otoacoustic emissions: A review. *Neurosci. Biobehav. Rev.* 18, 215–222. doi: 10.1016/0149-7634(94)90026-4

Müller, S., and Paulo, M. (2001). Transfer-function measurement with sweeps. *J. Audio Eng. Soc* 49, 443–471.

Neumann, J., Uppenkamp, S., and Kollmeier, B. (1994). Chirp evoked otoacoustic emissions. *Hear. Res.* 79, 17–25. doi: 10.1016/0378-5955(94)90123-6

Reavis, K. M., Phillips, D. S., Fausti, S. A., Gordon, J. S., Helt, W. J., Wilmington, D., et al. (2008). Factors affecting sensitivity of distortion-product otoacoustic emissions to ototoxic hearing loss. *Ear. Hear.* 29, 875–893. doi: 10.1097/AUD.0b013e318181ad99

Ren, T., He, W., Scott, M., and Nuttall, A. L. (2006). Group delay of acoustic emissions in the ear. *J. Neurophysiol.* 96, 2785–2791. doi: 10.1152/jn.00374.2006

Ren, T., Zheng, J., He, W., and Nuttall, A. L. (2013). Measurement of amplitude and delay of stimulus frequency otoacoustic emissions. *J. Otol.* 8, 57–62. doi: 10.1016/S1672-2930(13)50008-0

Ruggero, M. A. (2004). Comparison of group delays of 2f1− f2 distortion product otoacoustic emissions and cochlear travel times. *Acoust. Res. Lett. Online* 5, 143–147. doi: 10.1121/1.1771711

Shera, C. A., and Bergevin, C. (2012). Obtaining reliable phase-gradient delays from otoacoustic emission data. *J. Acoust. Soc. Am.* 132, 927–943. doi: 10.1121/1.4730916

Shera, C. A., and Guinan, J. J. (1999). Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs. *J. Acoust. Soc. Am.* 105, 782–798. doi: 10.1121/1.426948

Shera, C. A., and Guinan, J. J. (2003). Stimulus-frequency-emission group delay: A test of coherent reflection filtering and a window on cochlear tuning. *J. Acoust. Soc. Am.* 113, 2762–2772. doi: 10.1121/1.1557211

Shera, C. A., Guinan, J. J., and Oxenham, A. J. (2002). Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proc. Natl. Acad. Sci. U.S.A.* 99, 3318–3323. doi: 10.1073/pnas.032675099

Siegel, J. H., Cerka, A. J., Recio-Spinoso, A., Temchin, A. N., van Dijk, P., and Ruggero, M. A. (2005). Delays of stimulus-frequency otoacoustic emissions and cochlear vibrations contradict the theory of coherent reflection filtering. *J. Acoust. Soc. Am.* 118, 2434–2443. doi: 10.1121/1.2005867

Talmadge, C. L., Long, G. R., Tubis, A., and Dhar, S. (1999). Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions. *J. Acoust. Soc. Am.* 105, 275–292. doi: 10.1121/1.424584

Tognola, G., Grandori, F., and Ravazzani, P. (1997). Time-frequency distributions of click-evoked otoacoustic emissions. *Hear. Res.* 106, 112–122. doi: 10.1016/S0378-5955(97)00007-5

Wagner, W., Plinkert, P. K., Vonthein, R., and Plontke, S. K. (2008). Fine structure of distortion product otoacoustic emissions: Its dependence on age and hearing threshold and clinical implications. *Eur. Arch. Otorhinolaryngol.* 265, 1165–1172. doi: 10.1007/s00405-008-0593-0

Whitehead, M. L., Stagner, B. B., Lonsbury-Martin, B. L., and Martin, G. K. (1995). Effects of ear-canal standing waves on measurements of distortion-product otoacoustic emissions. *J. Acoust. Soc. Am.* 98, 3200–3214. doi: 10.1121/1.413810

Zweig, G., and Shera, C. A. (1995). The origin of periodicity in the spectrum of evoked otoacoustic emissions. *J. Acoust. Soc. Am.* 98, 2018–2047. doi: 10.1121/1.413320

frontiers | Frontiers in Neuroscience

TYPE Brief Research Report
PUBLISHED 25 October 2022
DOI 10.3389/fnins.2022.1031732

OPEN ACCESS

EDITED BY
Weidong Gao,
Beijing University of Posts
and Telecommunications (BUPT),
China

REVIEWED BY
Ruiquan Ge,
Hangzhou Dianzi University, China
Ruxin Wang,
Shenzhen Institutes of Advanced
Technology (CAS), China

*CORRESPONDENCE
Li Xu
shirleyxu@sjtu.edu.cn
Hong Liu
liuh@sumhs.edu.cn

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

RECEIVED 30 August 2022
ACCEPTED 06 October 2022
PUBLISHED 25 October 2022

CITATION
Xiong H, Chen H, Xu L, Liu H, Fan L,
Tang Q and Cho H (2022) A survey
of data element perspective:
Application of artificial intelligence
in health big data.
Front. Neurosci. 16:1031732.
doi: 10.3389/fnins.2022.1031732

A survey of data element perspective: Application of artificial intelligence in health big data

Honglin Xiong[1], Hongmin Chen[1], Li Xu[1]*, Hong Liu[2]*,
Lumin Fan[2,3], Qifeng Tang[4,5,6] and Hsunfang Cho[5,6]

[1]Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, China, [2]Business School, University of Shanghai for Science and Technology, Shanghai, China, [3]Operation Management Department, East Hospital Affiliated to Tongji University, Shanghai, China, [4]Department of Computer Science and Engineering, East China University of Science and Technology, Shanghai, China, [5]National Engineering Laboratory for Big Data Distribution and Exchange Technologies, Shanghai, China, [6]Shanghai Data Exchange Corporation, Shanghai, China

Artificial intelligence (AI) based on the perspective of data elements is widely used in the healthcare informatics domain. Large amounts of clinical data from electronic medical records (EMRs), electronic health records (EHRs), and electroencephalography records (EEGs) have been generated and collected at an unprecedented speed and scale. For instance, the new generation of wearable technologies enables easy-collecting peoples' daily health data such as blood pressure, blood glucose, and physiological data, as well as the application of EHRs documenting large amounts of patient data. The cost of acquiring and processing health big data is expected to reduce dramatically with the help of AI technologies and open-source big data platforms such as Hadoop and Spark. The application of AI technologies in health big data presents new opportunities to discover the relationship among living habits, sports, inheritances, diseases, symptoms, and drugs. Meanwhile, with the development of fast-growing AI technologies, many promising methodologies are proposed in the healthcare field recently. In this paper, we review and discuss the application of machine learning (ML) methods in health big data in two major aspects: (1) Special features of health big data including multimodal, incompletion, time validation, redundancy, and privacy. (2) ML methodologies in the healthcare field including classification, regression, clustering, and association. Furthermore, we review the recent progress and breakthroughs of automatic diagnosis in health big data and summarize the challenges, gaps, and opportunities to improve and advance automatic diagnosis in the health big data field.

KEYWORDS

healthcare big data, machine learning, automatic diagnosis, healthcare informatics, data elements, artificial intelligence

Introduction

The global "digital divide" *status quo* is quickly changing with the progress in artificial intelligence (AI) technologies and their application area expansion. Nowadays, AI has been widely researched and achieves great success in recent years, and the heart of AI technologies is machine learning (ML) algorithms (Xiong et al., 2022). With the development of the digital economy, Internet, Internet of things (IoT), mobile Internet, and cloud technologies, the application of AI based on health big data presents an explosive increase in recent years (Gokmen and Vlasov, 2016; Dolley, 2018; Ngiam and Khor, 2019; Ye et al., 2021; Weerasinghe et al., 2022). Besides, large amounts of personal health records (PHRs), electronic medical records (EMRs), and electronic health records (EHRs) in hospitals, many governments, and health organizations built the public health monitoring system to collect health data (Heart et al., 2017), such as NEDSS (National Electronic Disease Surveillance System) (National Electronic Disease Surveillance System Working Group, 2001), ProMED-mail (Yu and Madoff, 2004), GPHIN (Global Public Health Intelligence Network) (Dion et al., 2015), HealthMap (Freifeld et al., 2008), MediSys (Linge et al., 2010) and BioCaster (Collier et al., 2008). Among the public health regulatory systems, the representative system is NEDSS. It first defined the standard data protocol to ensure the medical or healthcare data with the identical data format collected across the country. Then, it enables large organizations to upload data automatically through electronic data interchange. The system mainly focused on the collection, exchange, and reporting of diseases and is lagging behind in knowledge mining and early disease warning. Meanwhile, the Internet giants like Google, Facebook, and Twitter collected large amounts of Internet social network data through their products and achieved influenza and other infectious diseases for early warning and tracking (Ginsberg et al., 2009; Signorini et al., 2011; Oflac et al., 2015). Google developed flu outbreak forecast software Google Flu, and the corresponding research result was published in Nature which invoked a large influence on the academic community (Oflac et al., 2015). However, recent research showed that the above-mentioned model in the prediction of flu outbreak existed big defects due to the instability of social network data (Lazer et al., 2014). Intel and IBM companies have also tried to use AI technologies for diabetes control (Nachman et al., 2010; Neuvirth et al., 2011) and the research results were published at the top conference of KDD (Knowledge Discovery in Database) (Neuvirth et al., 2011).

It is widely accepted that health big data have the potential to help physicians to improve diagnosis and aid drug usage. However, there exist many challenges in processing health big data even though researchers have achieved a lot of good results and applications. Except for five major features (5Vs) Volume, Velocity, Variety, Veracity, and Value, health big data have five additional special features (shown in **Figure 1**) as follows:

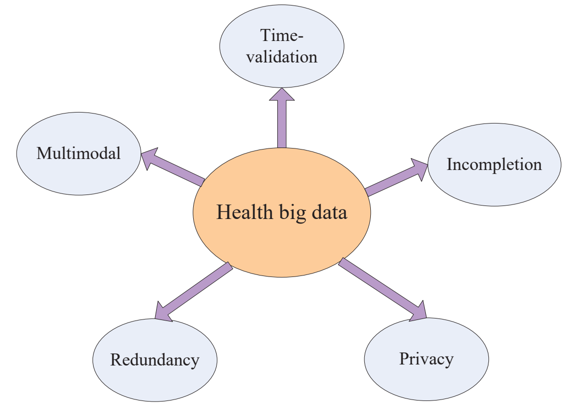


FIGURE 1
Five special features of health big data.

(1) Multimodal: healthcare data consist of text data, image data, and numerical data.

(2) Incompletion: There is a gap between medical data collection and treatment, which cause disease information reflection not enough. At the same time, recording data manually would have deviation, incompletion, and expression uncertainty due to subjective cognition.

(3) Time validation: There is progress between the patient's treatment and the disease outbreak. For example, electrocardiogram (ECG) and electroencephalogram (EEG) are time signals which have strong time-validated properties.

4) Redundancy: There are many same records stored in the healthcare data system. Take EHRs for example, physicians who serve in community hospitals often input multiple records due to unfamiliar computer operations, especially in China.

5) Privacy: It is inevitably related to the patient's private information when researchers process healthcare data. Disclosure of patients' privacy information will hurt patients' lives.

The rest of this paper is organized as follows. Section II introduces big data technologies like Hadoop, Spark, and Storm. The artificial intelligence technologies in health big data are described in section III. Section IV summarizes this paper.

Big data technologies

In the big data era, the traditional data management framework which is based on a relational database management system (RDBMS) is challenged by the increasing data deluge. The old framework is unable to deal with the growing amount of unstructured data. Therefore, new technologies were developed to serve the need for big data management, in the aspects of file systems and programming models. These technologies aim at providing scalability as well as fault tolerance, to handle the huge volume and heterogeneity of big data (Wang, 2017). Big data

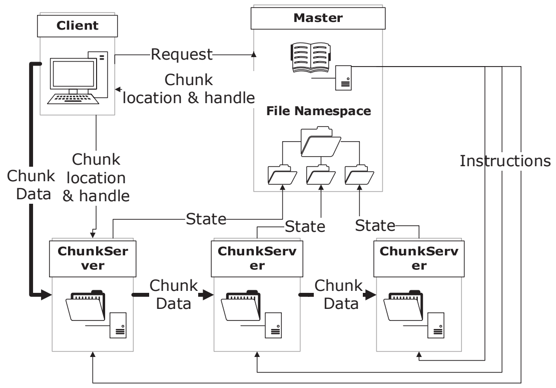


FIGURE 2
Google file system architecture.

have a wide range of applications, including Smart Grid cases, E-health, Internet of Things, Public utilities, Transportation and logistics, and other areas. The following passages introduce newly developed big data technologies in two aspects in detail.

Distributed file system

Although Moore's law promised that the storage capacity of computer chips doubles roughly every 18 months or so, current magnetic storage technology relies on a million atoms per bit, and the quantity of data grows much faster (Bradley, 2017). It is in great demand to develop an efficient and persistent distributed file system. In 2000, Brewer proposed the CAP theorem which states that it is impossible to meet the requirements of consistency, availability, and partition tolerance in an asynchronous distributed read/write system (Wang and Manzie, 2022). As frequent requests are common in big data scenarios, distributed file systems are commonly designed as AP systems, in which only eventual consistency rather than strong consistency is ensured.

As a pioneer in the attempts of providing users with high-performance services with a distributed file system, Google File System (GFS) achieved great success and its concepts were inherited by a lot of its successors. It features work division between control and storage servers, and replication of the same data, to provide performance and reliability in this way (Ghemawat et al., 2003). Its basic architecture and the data flow in it during a writing procedure was shown in **Figure 2**.

The highlight of this architecture includes that the control flow is separated from the data flow which leads to higher performance and that the replicas of data offer both reliability and efficiency under good management. Meanwhile, this system also has limitations in supporting small files, for its specific design purpose to support Google's own service.

Some successors of GFS are generally different implementations of the same idea, for example, HDFS (Kumari and Bucker, 2022) and KosmosFS. Others made some improvements to meet their own demands. Facebook developed Haystack which reduces disk operations for metadata lookups and increases overall throughput to support their Photos application (Beaver et al., 2010). Taobao developed TFS (Fu et al., 2014) which provides significantly higher performance in dealing with small files to support their online shopping service.

In conclusion, after many years of development, distributed file systems are relatively mature, and it is a prevailing trend to develop a customized DFS for a certain field.

MapReduce framework

As scalability and performance are two of the key requirements of a big data system, parallel computing must be implemented to offer these features. However, traditional parallel programming models fail in migrating to big data systems which consist of a massive number of servers over a wide area. In recent years, a lot of programming models were proposed to provide solutions to this specific need.

As the forerunner in distributing heavy computations across thousands of machines, MapReduce abstracted two basic operations from a broad variety of real-world tasks (Kalia and Gupta, 2021). The map function takes an input pair and produces a set of intermediate key/value pairs which will then be grouped and passed to the Reduce function. Reduce function is responsible for merging a set of values for one key to form a possibly smaller set of values. Once programmers give the proper definition to the two operations, the underlying runtime system will automatically parallelize and distribute the computation and handle other details including machine failures and inter-machine communication. The major part of its work procedure is illustrated in **Figure 3**.

Many programming models have been proposed afterward. Some provided considerable improvement to the MapReduce model. Microsoft developed Dryad (Isard et al., 2007) in which a job is abstracted as a directed acyclic graph. Each vertex is a program, and data channels are represented by edges. Higher generality is reached as data channels can be customized to support functions more than Map and Reduce. Spark (Solovyev et al., 2010) introduced an abstraction called resilient distributed datasets (RDDs) and parallel operations on them. An RDD represents a read-only collection of objects across a set of machines. By combining parallel operations based on data, Spark avoids redundant I/O operations and multiplied the performance. Other models focus on specific categories of distributed computing. Pregel (Malewicz et al., 2010) aims at large-scale graph processing, in which poor locality of memory access and very little work per vertex often lead to poor efficiency. Storm, as a stream processing model, offers outstanding performance in event processing and incremental computation.

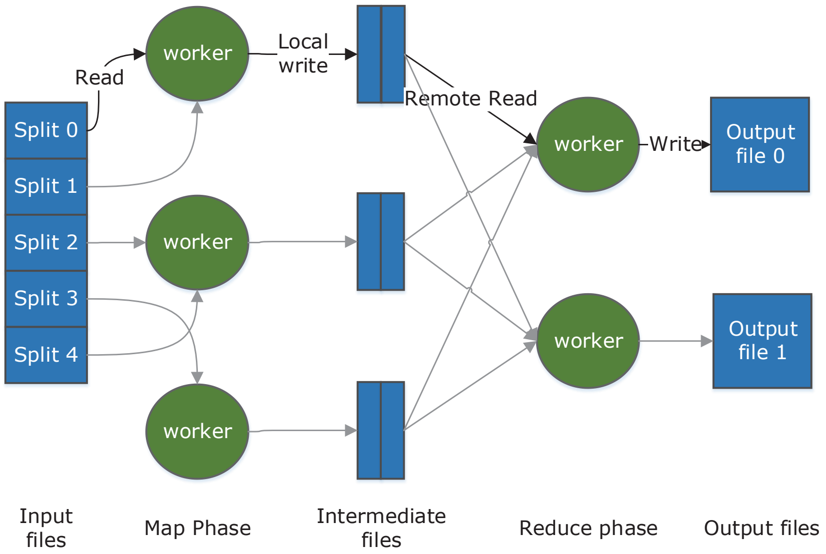


FIGURE 3
MapReduce work procedure.

Artificial intellectual technologies

In the big data era, various public hospitals and private healthcare providers are producing large amounts of data that are difficult to process. Therefore, powerful automatic artificial intellectual algorithms are needed for the analysis and processing of useful information from healthcare data. This information is very precious for healthcare specialists and physicians to apprehend the cause of diseases and for providing better and cost-effective treatment to patients. To improve prediction accuracy, there are various artificial intelligence technologies such as classification, regression, clustering, and association used in healthcare data analysis to increase the healthcare provider's capability for making the decision in regard to patients health. There are large amounts of research resources available regarding artificial intellectual application in health big data which are presented in subsequent sections with their advantages and disadvantages.

Classification

One of the data analysis tasks is classification, which divides data into target labels. Each data point is predicted into the target label by a pattern classifier. For instance, hypertension patients can be classified into three stages of stage 1 hypertension, stage 2 hypertension, and stage 3 hypertension (Wermelt and Schunkert, 2017) on the basis of a supervised classification model. Dataset is often partitioned into a training set, validation dataset, and testing dataset. The training dataset is utilized for training the classifier. The validation dataset is used to tune the classifier parameters to achieve optimal performance. Testing dataset verifies the classification accuracy. **Figure 4** shows the entire flowchart of classification.

In the ML domain, SVM as a supervised classifier is widely used for classification (Tsang et al., 2005). It is widely applied in healthcare data recently. Fei proposed the PSO-SVM model which has a strong global search capability (Fei, 2010), and the PSO-SVM model is applied to the diagnosis of arrhythmia cordis, in which PSO is used to determine the free parameters of the support vector machine (Cuong-Le et al., 2022). The testing results showed that the average classification accuracy is 95.65%. Huang et al. (2008) developed a hybrid SVM-based strategy with feature selection to render a diagnosis between breast cancer and fibroadenoma and to find the important risk factor for breast cancer (Azar and El-Said, 2014). The experimental results showed that the features {HSV-1, HHV-8} or {HSV-1, HHV-8, CMV} could achieve identical high accuracy, at 86% of the average overall hit rate. Zheng et al. (2014) used a hybrid of K-means and support vector machine (K-SVM) algorithms to extract useful information and diagnose the tumor. According to 10-fold cross-validation, the developed methodology which was tested on the Wisconsin Diagnostic Breast Cancer (WDBC) dataset from the University of California—Irvine ML repository, increased the accuracy to 97.38%. Avci utilized the genetic-support vector machine (GSVM) approach to classify the Doppler signals of heart valve diseases (Gonzalez-Abril et al., 2014). With the combination of feature extraction and classification from measured Doppler signal waveforms, the performance of the GSVM system showed that this GSVM system is effective to detect Doppler heart

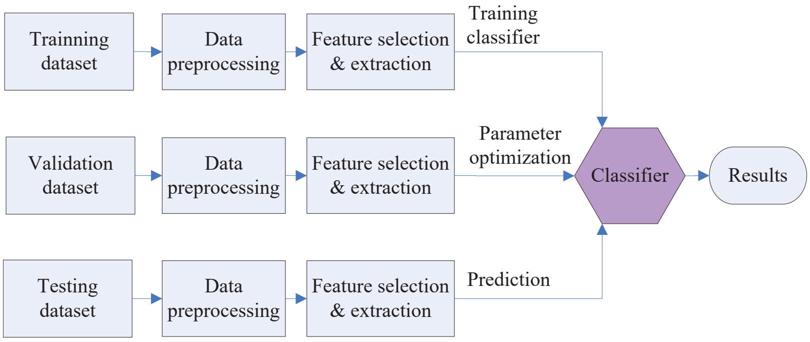


FIGURE 4
Classification flowchart.

sounds. The average rate of correct classification rate was about 95%.

A decision tree (DT) is a common ML method for constructing prediction models from data. The models are obtained by recursively partitioning the data space and fitting a simple prediction model within each partition (Kotsiantis, 2013; Loh, 2014). Due to its results with features of human-readable and interpretable, DT is widely used by many researchers in the healthcare field. Khan et al. (2008) proposed to investigate a hybrid scheme based on fuzzy decision trees, as an efficient alternative to predict breast cancer survivability for personalized healthcare. The experimental results showed that, for cancer prognosis, hybrid fuzz decision tree classification can achieve an average accuracy of 85%. Levashenko et al. (2016) proposed fuzzy decision trees in the medical decision-making support system. The classification accuracy of breast cancer was over 96%. Hassan et al. (2011) developed a decision tree with a CART classification algorithm to forecast response to therapy with 200 chronic hepatitis C patients. The overall classification error was 20%, and 80% was the best accuracy. Moon et al. (2012) developed decision tree models for characterizing smoking patterns in older adults. Their results suggest that social workers need to provide more customized and individualized interventions to older adults. Chang and Chen (2009) applied a decision tree and neural network to increase the quality of dermatologic diagnosis. Using sensitivity analysis combined with the decision tree model, on the contrary, has the least accuracy, which is 80.33%.

A neural network (NN) is based on a biological nervous system having multiple interrelated processing elements known as neurons, functioning in unity to solve a classification problem. Rules extracted from the trained model help to improve the interoperability of the learned network (Schmidhuber, 2015). Er et al. (2010) developed an artificial neural network (ANN) to diagnose chest diseases. Sokolov (2018) presents recent research on approaches in autonomous systems for combining multiple modalities for emotion estimation based on neural networks. Sharma and Parmar (2020) utilized a neural network approach to analyze a heart disease dataset the experimental results proved better accuracy (90.76%) than other optimizations. It is applied to heart disease datasets and finds out a good prediction (Sharma and). In the past several years, intricate neural networks have inspired the further development of intelligent systems. Many disciplines, including the complex field of medicine, neuroimaging modalities, and diagnosis of the disease, have taken advantage of the useful applications of artificial neural networks (Yang et al., 2018; Deperlioglu et al., 2020).

Bayesian decision theory is a basic method under the statistic framework, and it is extended easily to do classification tasks (Chickering et al., 2004). Liu and Lu (2009) proposed to use Bayesian belief network (BBN) as decision support for the higher-level risk estimate which can represent the probabilistic relationships between all kinds of health effects and air pollutants. Dawson et al. (2015) used the Bayesian network to produce the baseline distribution by taking the joint distribution of the data and conditioning it on attributes that are responsible for anomaly pattern detection for disease outbreaks. Curiac et al. (2009) analyzed the psychiatric patient data using BBN in making a significant decision regarding patient health suffering from psychiatric disease and performed an experiment on real data obtained from Lugoj Municipal Hospital.

Long et al. (2015) proposed a heart disease diagnosis system using rough set-based attribute reduction and interval type-2 fuzzy logic system (IT2FLS). The experimental results showed that it could efficiently find minimal attribute reduction from the high-dimensional dataset that enhances the performance of the classification system. The use of an interval type-2 fuzzy logic system for the classification of heart disease datasets to handle the uncertainties and noisiness of these datasets was successful. Nahar et al. (2013) presented the potential of an expert judgment-based (i.e., medical knowledge-driven) feature selection process (termed as MFS). The medical knowledge-based feature selection method has shown promise for use in heart disease diagnostics. The main classification methods used in healthcare big data are shown in **Table 1**.

The application of classification analysis methods in medicine is getting more and more advanced, not only is it used extensively in disease diagnosis, but also there will be more breakthroughs in disease treatment options in future, and all of these expectations become more apparent in the near future.

Regression

Regression analysis is a statistical method to determine the quantitative relationship between two or more variables. Based on observational data, regression analysis could establish appropriate dependencies between variables and analyze the inherent rules of data (Merrick et al., 2022). It is widely used for forecasting in the healthcare field. Gutiérrez et al. (2010) proposed a hybrid multi-logistic methodology, named logistic regression using initial and radial basis function (RBF) covariates. Agarwal (2011). developed weighted support vector regression (SVR) approach for remote healthcare monitoring. Vinsnes et al. (2001) developed a regression analysis approach for healthcare personnel's attitudes to predict nursing assistants' attitudes. Ko and Osei-Bryson (2004) explored the productivity impact of information technology (IT) in the healthcare industry using a regression spline (RS)-based approach. Luo et al. (2012) presented scalable orthogonal regression (SOR) for non-redundant feature selection and its healthcare applications.

Regression analysis can accurately measure the degree of correlation between factors and the degree of the regression fit to improve the effectiveness of prediction, which is of great significance in medical diagnosis. More recently, regression analysis is one of the most frequently used analytical techniques in disease diagnosis and etiology analysis (Hannan et al., 2010; Liu et al., 2018; Jfri et al., 2021).

Clustering

Clustering is an unsupervised learning method that is different from classification. Clustering is a process of classifying data into different classes or clusters, so objects in the same cluster have a large similarity, and objects between different clusters have a large degree of dissimilarity (Caron et al., 2018). Clustering is also used in the healthcare field. Sinaga and Yang (2020) proposed a novel unsupervised k-means (U-k-means) clustering algorithm which automatically finds an optimal number of clusters without giving any initialization and parameter selection. Stein et al. (2007) used data clustering techniques to develop health state descriptions based on data from 66 women who completed the EORTC QLQ-C30 over a 6-month period while receiving chemotherapy for ovarian cancer. Belciug et al. (2010) detected breast cancer recurrence with the help of a clustering-based approach. Zhao et al. (2020) proposed to propose a new deep learning and clustering UDFCMN (Unsupervised Deep Fuzzy C-Means clustering Network) model, to cluster lung cancer patients from lung CT images; these results also indicate that this method has practical applications in lung cancer pathogenesis studies and provide useful guidelines for personalized cancer therapy. Balasubramanian and Umarani (2012) analyzed the impact of fluoride on human health (dental) with the help of a clustering-based method and found meaningful hidden patterns which gave meaningful decision-making to this socio-economic real-world health hazard. In addition, some researchers have also used clustering methods to early detect Alzheimer's disease (Escudero et al., 2011; Holilah et al., 2021). The main clustering methods used in healthcare big data are shown in **Table 2**.

Cluster analysis is essentially finding a statistic that objectively reflects the affinity of an element and then classifying the elements into categories based on this statistic. Cluster analysis decomposes the symptoms of chronic diseases and is used to assess the quality of life in chronic diseases, such as lung cancer; cluster analysis is very effective in assessing these diseases.

Association

Association is one of the most vital approaches to data mining that is used to find out the frequent patterns, and interesting relationships among a set of data items in the data repository. Frequent patterns are patterns that appear frequently in a dataset. The initial motivations of the association rules were raised for the issue of the market basket analysis. The association process analyzes the customer's shopping habits by discovering the association between the different items placed in the "shopping basket" by the customer. The discovery of this association can help retailers understand which goods are frequently purchased by customers at the same time, so as to help them develop better marketing strategies (Tomar and Agarwal, 2013). Association also has a great impact in the healthcare field to detect the relationships among diseases, health status, and symptoms. Nahar et al. (2013) presented a rule extraction experiment on heart disease data using different rule-mining algorithms (*Apriori*, Predictive *Apriori*, and Tertius). Further rule-mining-based analysis was undertaken by categorizing data based on gender, and significant risk factors for heart disease were found for both men and women. Ji et al. (2010) developed a new interestingness measure, exclusive causal-leverage, based on an experience-based fuzzy recognition-primed decision (RPD) model. On the basis of this new measure, a new association rule algorithm is proposed to discover infrequent causal relationships in electronic health databases (Horton et al., 2019). In addition, Soni and Vyas (2010) used the associative method to construct a classifier for predictive analysis in healthcare data mining.

TABLE 1 Main application of classification methods in healthcare big data.

Method	Scenes	Features
SVM	Diagnosis of arrhythmia cordis; diagnosis between breast cancer and fibroadenoma; diagnosis of the tumor; detect Doppler heart sounds and so on	Non-linear mapping; low generalization error rate, fast classification; suitable for small samples, excellent generalization ability, etc.,
DT	Predict breast cancer survivability; medical decision-making support system; characterizing smoking patterns and so on	Simple to understand, easy to explain, visualization, and wide applicability; prone to overfitting, in addition, small changes in the data can affect the results and are unstable
NN	Including the complex field of medicine, neuroimaging modalities, and diagnosis of the disease; image analysis and interpretation	With self-learning function; no *a priori* assumptions about the problem model are required. suitable for some problems with very complex environmental information, unclear knowledge background, and unclear inference rules.
BN	Anomaly pattern detection for disease outbreaks; regarding patient health suffering from psychiatric disease	Distribution of input data in each layer of the network is relatively stable, which accelerates the model learning speed; makes the model less sensitive to the parameters in the network, simplifies the tuning process, and makes the network learning more stable

TABLE 2 Main application of clustering methods in healthcare big data.

Method	Scenes	Features
k-means	Health state descriptions; Alzheimer's disease; health hazards and so on	Fast convergence; better clustering effect; stronger interpretability of the model and so on
Fuzzy C-means	Lung cancer patients from lung CT images	Clustering objectively and accurately

Using the association analysis method to discover the relationship between the attributes in the medical dataset, especially some general factors such as age and smoking habits, and the measured body organ function indices related to the possibility of disease, the doctor can accurately determine the possibility of disease through the patient's characteristics, which is very meaningful for medical diagnosis, and the future application of the association analysis method to predict diseases and develop treatment plans based on vital signs.

In summary, AI as a role in healthcare big data, its effects on the development of the medical industry, applications of AI in medicine, challenges, and promises of both AI and big data with respect to healthcare, and prevailing techniques (methods such as deep neural network, convolutional neural network, and recurrent neural network) and tools for performance optimization of healthcare big data can be used by the medical industry.

Conclusion

This paper investigated the application of AI technologies in health big data based on a data elements perspective. Traditional data management framework, which is based on a relational database management system (RDBMS), is hard to deal with a growing amount of healthcare data. Big data processing frameworks like Hadoop (Spark et al.) are employed in the data preprocess stage to accommodate big data and accelerate computing efficiency. We found that there is no single ML method that gives consistently good results for all kinds of health big data. The performance of ML methods depends on the type of dataset that researchers have taken for doing the experiment. To get the higher performance of ML method, most of the research combined many artificial methods to complement the deficiency of each one called hybrid method or integrated method or assemble method.

In addition, it is well known in the AI field that feature selection and extraction are very important factors that affect the performance of artificial methods. Features are extracted and selected on the basis of healthcare/medical domain knowledge and optimal techniques in normal conditions. For healthcare providers and medical providers, AI technologies are widely utilized to make effective decisions in regard to how to enhance patients' health, how to provide healthcare service at low cost, and how to remind physicians to avoid misusing drugs and misdiagnosing.

Data availability statement

The original contributions presented in this study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

Author contributions

HX analyzed the methods and penned the manuscript. HC, LX, and HL gave important suggestions. QT, LF, and HFC participated in discussions and provided some literature resources. HC revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the Major Research Project of Philosophy and Social Science of the Ministry of Education of PRC (Grant No. 20JZD010).

Acknowledgments

The authors would like to thank Chen and Xu who give their precious suggestions on improving manuscript quality.

Conflict of interest

Authors QT and HFC were employed by Shanghai Data Exchange Corporation.

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

Publisher's note

References

Agarwal, S. (2011). "Weighted support vector regression approach for remote healthcare monitoring," in *Proceedings of the 2011 international conference on recent trends in information technology (ICRTIT)*, (Chennai: IEEE), 969–974. doi: 10.1109/ICRTIT.2011.5972437

Azar, A. T., and El-Said, S. A. (2014). Performance analysis of support vector machines classifiers in breast cancer mammography recognition. *Neural Comput. Appl.* 24, 1163–1177. doi: 10.1007/s00521-012-1324-4

Balasubramanian, T., and Umarani, R. (2012). "An analysis on the impact of fluoride in human health (dental) using clustering data mining technique," in *Proceedings of the international conference on pattern recognition, informatics and medical engineering*, (Salem: IEEE), 370–375. doi: 10.1109/ICPRIME.2012.6208374

Beaver, D., Kumar, S., Li, H. C., Sobel, J., and Vajgel, P. (2010). "Finding a needle in haystack: Facebook's photo storage," in *Proceedings of the 9th USENIX symposium on operating systems design and implementation (OSDI 10)*, (Vancouver: IEEE), 1–8.

Belciug, S., Salem, A. B., Gorunescu, F., and Gorunescu, M. (2010). "Clustering-based approach for detecting breast cancer recurrence," in *Proceedings of the 2010 10th international conference on intelligent systems design and applications*, (Cairo: IEEE), 533–538. doi: 10.1109/ISDA.2010.5687211

Bradley, D. (2017). Single-atom memory maintains Moore's Law. *Materialstoday* 20:225. doi: 10.1016/j.mattod.2017.04.021

Caron, M., Bojanowski, P., Joulin, A., and Douze, M. (2018). "Deep clustering for unsupervised learning of visual features," in *Proceedings of the European conference on computer vision (ECCV)*, (Tel Aviv: IEEE), 132–149.

Chang, C. L., and Chen, C. H. (2009). Applying decision tree and neural network to increase quality of dermatologic diagnosis. *Expert Syst. Appl.* 36, 4035–4041. doi: 10.1016/j.eswa.2008.03.007

Chickering, M., Heckerman, D., and Meek, C. (2004). Large-sample learning of Bayesian networks is NP-hard. *J. Mach. Learn. Res.* 5, 1287–1330.

Collier, N., Doan, S., Kawazoe, A., Goodwin, R. M., Conway, M., Tateno, Y., et al. (2008). BioCaster: Detecting public health rumors with a Web-based text mining system. *Bioinformatics* 24, 2940–2941. doi: 10.1093/bioinformatics/btn534

Cuong-Le, T., Nghia-Nguyen, T., Khatir, S., Trong-Nguyen, P., Mirjalili, S., and Nguyen, K. D. (2022). An efficient approach for damage identification based on improved machine learning using PSO-SVM. *Eng. Comput.* 38, 3069–3084. doi: 10.1007/s00366-021-01299-6

Curiac, D. I., Vasile, G., Banias, O., Volosencu, C., and Albu, A. (2009). "Bayesian network model for diagnosis of psychiatric diseases," in *Proceedings of the ITI 2009 31st international conference on information technology interfaces*, (Cavtat: IEEE), 61–66. doi: 10.1109/ITI.2009.5196055

Dawson, P., Gailis, R., and Meehan, A. (2015). Detecting disease outbreaks using a combined Bayesian network and particle filter approach. *J. Theor. Biol.* 370, 171–183. doi: 10.1016/j.jtbi.2015.01.023

Deperlioglu, O., Kose, U., Gupta, D., Khanna, A., and Sangaiah, A. K. (2020). Diagnosis of heart diseases by a secure internet of health things system based on autoencoder deep neural network. *Comput. Commun.* 162, 31–50. doi: 10.1016/j.comcom.2020.08.011

Dion, M., AbdelMalik, P., and Mawudeku, A. (2015). Big data: Big data and the global public health intelligence network (GPHIN). *Can. Commun. Dis. Rep.* 41, 209–214. doi: 10.14745/ccdr.v41i09a02

Dolley, S. (2018). Big data's role in precision public health. *Front. Public Health* 68:6. doi: 10.3389/fpubh.2018.0006

Er, O., Yumusak, N., and Temurtas, F. (2010). Chest diseases diagnosis using artificial neural networks. *Expert Syst. Appl.* 37, 7648–7655. doi: 10.1016/j.eswa.2010.04.078

Escudero, J., Zajicek, J. P., and Ifeachor, E. (2011). Early detection and characterization of Alzheimer's disease in clinical scenarios using Bioprofile concepts and K-means. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2011, 6470–6473. doi: 10.1109/IEMBS.2011.6091597

Fei, S. W. (2010). Diagnostic study on arrhythmia cordis based on particle swarm optimization-based support vector machine. *Expert Syst. Appl.* 37, 6748–6752. doi: 10.1016/j.eswa.2010.02.126

Freifeld, C. C., Mandl, K. D., Reis, B. Y., and Brownstein, J. S. (2008). HealthMap: Global infectious disease monitoring through automated classification and visualization of Internet media reports. *J. Am. Med. Inf. Assoc.* 15, 150–157. doi: 10.1197/jamia.M2544

Fu, S., He, L., Huang, C., Liao, X., and Li, K. (2014). Performance optimization for managing massive numbers of small files in distributed file systems. *IEEE Trans. Parallel Distrib. Syst.* 26, 3433–3448. doi: 10.1109/TPDS.2014.2377720

Ghemawat, S., Gobioff, H., and Leung, S. T. (2003). "The Google file system," in *Proceedings of the nineteenth ACM symposium on operating systems principles*, New York, NY, 29–43. doi: 10.1145/945445.945450

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., and Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature* 457, 1012–1014.

Gokmen, T., and Vlasov, Y. (2016). Acceleration of deep neural network training with resistive cross-point devices: Design considerations. *Front. Neurosci.* 10:333. doi: 10.3389/fnins.2016.00333

Gonzalez-Abril, L., Nunez, H., Angulo, C., and Velasco, F. (2014). GSVM: An SVM for handling imbalanced accuracy between classes inbi-classification problems. *Appl. Soft Comput.* 17, 23–31. doi: 10.1016/j.asoc.2013.12.013

Gutiérrez, P. A., Hervás-Martínez, C., and Martínez-Estudillo, F. J. (2010). Logistic regression by means of evolutionary radial basis function neural networks. *IEEE Trans. Neural Netw.* 22, 246–263. doi: 10.1109/TNN.2010.2093537

Hannan, S. A., Manza, R. R., and Ramteke, R. J. (2010). Generalized regression neural network and radial basis function for heart disease diagnosis. *Int. J. Comput. Appl.* 7, 7–13. doi: 10.5120/1325-1799

Hassan, M., Abdalla, M. I., Ahmed, S. R., Akil, W., Esmat, G., Khamis, S., et al. (2011). The decision tree mode for prediction the response to the treatment in patients with chronic hepatitis C. *N. Y. Sci. J.* 4, 69–79.

Heart, T., Ben-Assuli, O., and Shabtai, I. (2017). A review of PHR, EMR and EHR integration: A more personalized healthcare and public health policy. *Health Policy Technol.* 6, 20–25. doi: 10.1016/j.hlpt.2016.08.002

Holilah, D., Bustamam, A., and Sarwinda, D. (2021). Detection of Alzheimer's disease with segmentation approach using K-Means Clustering and Watershed Method of MRI image. *J. Phys. Conf. Ser.* 1725:012009. doi: 10.1088/1742-6596/1725/1/012009

Horton, D. B., Bhullar, H., Carty, L., Cunningham, F., Ogdie, A., Sultana, J., et al. (2019). "Electronic health record databases," in *Pharmacoepidemiology*, eds B. L. Strom, S. E. Kimmel, and S. Hennessy (Hoboken, NJ: Wiley), 241–289. doi: 10.1002/9781119413431.ch13

Huang, C. L., Liao, H. C., and Chen, M. C. (2008). Prediction model building and feature selection with support vector machines in breast cancer diagnosis. *Expert Syst. Appl.* 34, 578–587. doi: 10.1016/j.eswa.2006.09.041

Isard, M., Budiu, M., Yu, Y., Birrell, A., and Fetterly, D. (2007). "Dryad: Distributed data-parallel programs from sequential building blocks," in *Proceedings of the 2nd ACM SIGOPS/eurosys European conference on computer systems*, Vol. 2007, (New York, NY), 59–72. doi: 10.1145/1272996.1273005

Jfri, A., Nassim, D., O'Brien, E., Gulliver, W., Nikolakis, G., and Zouboulis, C. C. (2021). Prevalence of hidradenitis suppurativa: A systematic review and meta-regression analysis. *JAMA Dermatol.* 157, 924–931. doi: 10.1001/jamadermatol.2021.1677

Ji, Y., Ying, H., Dews, P., Farber, M. S., Mansour, A., Tran, J., et al. (2010). "A fuzzy recognition-primed decision model-based causal association mining algorithm for detecting adverse drug reactions in postmarketing surveillance," in *Proceedings of the international conference on fuzzy systems*, (Barcelona: IEEE), 1–8. doi: 10.1109/FUZZY.2010.5584288

Kalia, K., and Gupta, N. (2021). Analysis of hadoop MapReduce scheduling in heterogeneous environment. *Ain Shams Eng. J.* 12, 1101–1110. doi: 10.1016/j.asej.2020.06.009

Khan, M. U., Choi, J. P., Shin, H., and Kim, M. (2008). Predicting breast cancer survivability using fuzzy decision trees for personalized healthcare. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2008, 5148–5151. doi: 10.1109/IEMBS.2008.4650373

Ko, M., and Osei-Bryson, K. M. (2004). The productivity impact of information technology in the healthcare industry: An empirical study using a regression spline-based approach. *Inf. Softw. Technol.* 46, 65–73. doi: 10.1016/S0950-5849(03)00110-1

Kotsiantis, S. B. (2013). Decision trees: A recent overview. *Artif. Intell. Rev.* 39, 261–283. doi: 10.1007/s10462-011-9272-4

Kumari, P. S., and Bucker, N. A. B. A. (2022). Data integrity verification using HDFS framework in data flow material environment using cloud computing. *Mater. Today Proc.* 60, 1329–1333. doi: 10.1016/j.matpr.2021.09.435

Lazer, D., Kennedy, R., King, G., and Vespignani, A. (2014). The parable of Google Flu: Traps in big data analysis. *Science* 343, 1203–1205. doi: 10.1126/science.1248506

Levashenko, V., Zaitseva, E., Kvassay, M., and Deserno, T. M. (2016). "Reliability estimation of healthcare systems using fuzzy decision trees," in *Proceedings of the 2016 federated conference on computer science and information systems (FedCSIS)*, (Gdansk: IEEE), 331–340. doi: 10.15439/2016F150

Linge, J. P., Steinberger, R., Fuart, F., Bucci, S., Belyaeva, J., Gemo, M., et al. (2010). "MedISys: Medical information system," in *Advanced ICTs for disaster management and threat detection: Collaborative and distributed frameworks*, eds E. Asimakopoulou and N. Bessis (Hershey, PA: IGI Global), 131–142.

Liu, K. F., and Lu, C. F. (2009). "BBN-based decision support for health risk analysis," in *Proceedings of the 2009 fifth international joint conference on INC, IMS and IDC*, Washington, DC, 696–702. doi: 10.1109/NCM.2009.187

Liu, M., Zhang, J., Adeli, E., and Shen, D. (2018). Joint classification and regression via deep multi-task multi-channel learning for Alzheimer's disease diagnosis. *IEEE Trans. Biomed. Eng.* 66, 1195–1206. doi: 10.1109/TBME.2018.2869989

Loh, W. Y. (2014). Fifty years of classification and regression trees. *Int. Stat. Rev.* 82, 329–348. doi: 10.1111/insr.12016

Long, N. C., Meesad, P., and Unger, H. (2015). A highly accurate firefly based algorithm for heart disease prediction. *Expert Syst. Appl.* 42, 8221–8231. doi: 10.1016/j.eswa.2015.06.024

Luo, D., Wang, F., Sun, J., Markatou, M., Hu, J., and Ebadollahi, S. (2012). "Sor: Scalable orthogonal regression for non-redundant feature selection and its healthcare applications," in *Proceedings of the 2012 SIAM international conference on data mining*, Alexandria, 576–587. doi: 10.1137/1.9781611972825.50

Malewicz, G., Austern, M. H., Bik, A. J., Dehnert, J. C., Horn, I., Leiser, N., et al. (2010). "Pregel: A system for large-scale graph processing," in *Proceedings of the 2010 ACM SIGMOD international conference on management of data*, (New York, NY: ACM), 135–146. doi: 10.1145/1807167.1807184

Merrick, L. F., Lozada, D. N., Chen, X., and Carter, A. H. (2022). Classification and regression models for genomic selection of skewed phenotypes: A case for disease resistance in winter wheat (*Triticum aestivum* L). *Front. Genet.* 13:835781. doi: 10.3389/fgene.2022.835781

Moon, S. S., Kang, S. Y., Jitpitaklert, W., and Kim, S. B. (2012). Decision tree models for characterizing smoking patterns of older adults. *Expert Syst. Appl.* 39, 445–451. doi: 10.1016/j.eswa.2011.07.035

Nachman, L., Baxi, A., Bhattacharya, S., Darera, V., Deshpande, P., Kodalapura, N., et al. (2010). "Jog falls: A pervasive healthcare platform for diabetes management," in *International conference on pervasive computing*, eds P. Floréen, A. Krüger, and M. Spasojevic (Berlin: Springer), 94–111.

Nahar, J., Imam, T., Tickle, K. S., and Chen, Y. P. P. (2013). Association rule mining to detect factors which contribute to heart disease in males and females. *Expert Syst. Appl.* 40, 1086–1093. doi: 10.1016/j.eswa.2012.08.028

National Electronic Disease Surveillance System Working Group (2001). National Electronic Disease Surveillance System (NEDSS): A standards-based approach to connect public health and clinical medicine. *J. Public Health Manag. Pract.* 43–50.

Neuvirth, H., Ozery-Flato, M., Hu, J., Laserson, J., Kohn, M. S., Ebadollahi, S., et al. (2011). "Toward personalized care management of patients at risk: The diabetes case study," in *Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining*, Vol. 8, San Diego, CA, 395–403.

Ngiam, K. Y., and Khor, W. (2019). Big data and machine learning algorithms for health-care delivery. *Lancet Oncol.* 20, e262–e273. doi: 10.1016/S1470-2045(19)30149-4

Oflac, B. S., Dobrucalı, B., Yavas, T., and Escobar, M. G. (2015). Services marketing mix efforts of a global services brand: The case of DHL Logistics. *Procedia Econ. Finance* 23, 1079–1083. doi: 10.1016/S2212-5671(15)00457-8

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Netw.* 61, 85–117. doi: 10.1016/j.neunet.2014.09.003

Sharma, S., and Parmar, M. (2020). Heart diseases prediction using deep learning neural network model. *Int. J. Innov. Technol. Explor. Eng.* 9, 124–137. doi: 10.35940/ijitee.C9009.019320

Signorini, A., Segre, A. M., and Polgreen, P. M. (2011). The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. *PLoS One* 6:e19467. doi: 10.1371/journal.pone.0019467

Sinaga, K. P., and Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access* 8, 80716–80727. doi: 10.1109/ACCESS.2020.2988796

Sokolov, S. (2018). Neural network based multimodal emotion estimation. *ICAS* 2018, 4–7.

Solovyev, A., Mikheev, M., Zhou, L., Dutta-Moscato, J., Ziraldo, C., An, G., et al. (2010). "SPARK: A framework for multi-scale agent-based biomedical modeling," in *Proceedings of the 2010 spring simulation multiconference*, Orlando, FL, 1–7. doi: 10.1145/1878537.1878541

Soni, S., and Vyas, O. P. (2010). Using associative classifiers for predictive analysis in health care data mining. *Int. J. Comput. Appl.* 4, 33–37.

Stein, K., Sugar, C., Velikova, G., and Stark, D. (2007). Putting the 'Q' in quality adjusted life years (QALYs) for advanced ovarian cancer–An approach using data clustering methods and the internet. *Eur. J. Cancer* 43, 104–113. doi: 10.1016/j.ejca.2006.09.007

Tomar, D., and Agarwal, S. (2013). A survey on data mining approaches for healthcare. *Int. J. Biosci. Biotechnol.* 5, 241–266.

Tsang, I. W., Kwok, J. T., Cheung, P. M., and Cristianini, N. (2005). Core vector machines: Fast SVM training on very large data sets. *J. Mach. Learn. Res.* 6, 363–392.

Vinsnes, A. G., Harkless, G. E., Haltbakk, J., Bohm, J., and Hunskaar, S. (2001). Healthcare personnel's attitudes towards patients with urinary incontinence INFORMATION POINT: Regression analysis. *J. Clin. Nurs.* 10, 455–462. doi: 10.1046/j.1365-2702.2001.00513.x

Wang, L. (2017). Heterogeneous data and big data analytics. *Automat. Control Inf. Sci.* 3, 8–15. doi: 10.12691/acis-3-1-3

Wang, Y., and Manzie, C. (2022). Robust distributed model predictive control of linear systems: Analysis and synthesis. *Automatica* 137:110141. doi: 10.1016/j.automatica.2021.110141

Weerasinghe, K., Scahill, S. L., Pauleen, D. J., and Taskin, N. (2022). Big data analytics for clinical decision-making: Understanding health sector perceptions of policy and practice. *Technol. Forecast. Soc. Change* 174:121222. doi: 10.1016/j.techfore.2021.121222

Wermelt, J. A., and Schunkert, H. (2017). Management of arterial hypertension. *Herz* 42, 515–526. doi: 10.1007/s00059-017-4574-1

Xiong, H., Fan, C., Chen, H., Yang, Y., Antwi, C. O., and Fan, X. (2022). A novel approach to air passenger index prediction: Based on mutual information principle and support vector regression blended model. *SAGE Open* 12:21582440211071102. doi: 10.1177/21582440211071102

Yang, Z., Huang, Y., Jiang, Y., Sun, Y., Zhang, Y. J., and Luo, P. (2018). Clinical assistant diagnosis for electronic medical record based on convolutional neural network. *Sci. Rep.* 8:6329. doi: 10.1038/s41598-018-24389-w

Ye, Y., Shi, J., Zhu, D., Su, L., Huang, J., and Huang, Y. (2021). Management of medical and health big data based on integrated learning-based health care system: A review and comparative analysis. *Comput. Methods Prog. Biomed.* 209:106293. doi: 10.1016/j.cmpb.2021.106293

Yu, V. L., and Madoff, L. C. (2004). ProMED-mail: An early warning system for emerging diseases. *Clin. Infect. Dis.* 39, 227–232. doi: 10.1086/422003

Zhao, Z., Zhao, J., Song, K., Hussain, A., Du, Q., Dong, Y., et al. (2020). Joint DBN and Fuzzy C-Means unsupervised deep clustering for lung cancer patient stratification. *Eng. Appl. Artif. Intell.* 91:103571. doi: 10.1016/j.engappai.2020.103571

Zheng, B., Yoon, S. W., and Lam, S. S. (2014). Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. *Expert Syst. Appl.* 41, 1476–1482. doi: 10.1016/j.eswa.2013.08.044

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 22 February 2023
DOI 10.3389/fnins.2023.1018037

Ascertaining the optimal myoelectric signal recording duration for pattern recognition based prostheses control

Mojisola Grace Asogbon[1], Oluwarotimi Williams Samuel[1]*, Ejay Nsugbe[2], Yongcheng Li[1], Frank Kulwa[1], Deogratias Mzurikwao[3], Shixiong Chen[1] and Guanglin Li[1]*

[1]CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology (SIAT), Chinese Academy of Sciences (CAS), Shenzhen Institute of Artificial Intelligence and Robotics for Society, Shenzhen, China, [2]Nsugbe Research Labs, Swindon, United Kingdom, [3]Unit of Biomedical Engineering, Department of Physiology, School of Engineering, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania

EDITED BY
Xiaomao Fan,
Shenzhen Technology University, China

REVIEWED BY
Alexander E. Olsson,
Faculty of Engineering, Lund University, Sweden
Juan Manuel Fontana,
National University of Río Cuarto, Argentina

*CORRESPONDENCE
Oluwarotimi Williams Samuel
✉ samuel@siat.ac.cn
Guanglin Li
✉ gl.li@siat.ac.cn

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

RECEIVED 12 August 2022
ACCEPTED 27 January 2023
PUBLISHED 22 February 2023

CITATION
Asogbon MG, Samuel OW, Nsugbe E, Li Y, Kulwa F, Mzurikwao D, Chen S and Li G (2023) Ascertaining the optimal myoelectric signal recording duration for pattern recognition based prostheses control.
Front. Neurosci. 17:1018037.
doi: 10.3389/fnins.2023.1018037

Introduction: Electromyogram-based pattern recognition (EMG-PR) has been widely considered an essentially intuitive control method for multifunctional upper limb prostheses. A crucial aspect of the scheme is the EMG signal recording duration (SRD) from which requisite motor tasks are characterized per time, impacting the system's overall performance. For instance, lengthy SRD inevitably introduces fatigue (that alters the muscle contraction patterns of specific limb motions) and may incur high computational costs in building the motion intent decoder, resulting in inadequate prosthetic control and controller delay in practical usage. Conversely, relatively shorter SRD may lead to reduced data collection durations that, among other advantages, allow for more convenient prosthesis recalibration protocols. Therefore, determining the optimal SRD required to characterize limb motion intents adequately that will aid intuitive PR-based control remains an open research question.

Method: This study systematically investigated the impact and generalizability of varying lengths of myoelectric SRD on the characterization of multiple classes of finger gestures. The investigation involved characterizing fifteen classes of finger gestures performed by eight normally limb subjects using various groups of EMG SRD including 1, 5, 10, 15, and 20 s. Two different training strategies including Between SRD and Within-SRD were implemented across three popular machine learning classifiers and three time-domain features to investigate the impact of SRD on EMG-PR motion intent decoder.

Result: The between-SRD strategy results which is a reflection of the practical scenario showed that an SRD greater than 5 s but less than or equal to 10 s (>5 and < = 10 s) would be required to achieve decent average finger gesture decoding accuracy for all feature-classifier combinations. Notably, lengthier SRD would incur more acquisition and implementation time and vice-versa. In inclusion, the study's findings provide insight and guidance into selecting appropriate SRD that would aid inadequate characterization of multiple classes of limb motion tasks in PR-based control schemes for multifunctional prostheses.

KEYWORDS

electromyogram (EMG), finger gestures, pattern recognition (PR), prostheses, signal recording duration (SRD)

1. Introduction

Upper limb loss precludes amputees from full exploration of their environment especially in accomplishing tasks that require their arm functions (Cordella et al., 2016; Wheaton, 2017). The varied setbacks faced by amputees during daily life activities have spurred the development of intelligent prosthetic limbs meant to intuitively restore their limb functions. At the forefront of this technology are myoelectric pattern recognition (PR) based prostheses that use decoded motion intent from surface electromyogram (EMG) signals for their control (Cordella et al., 2016; Vujaklija et al., 2016; Parajuli et al., 2019). In an archetypal PR-based prosthetic control pipeline, EMG signals of coordinated muscle activities of specific limb motion are recorded, processed, and motor tasks are decoded *via* machine learning algorithms which serve as control inputs to the device (Li et al., 2010; Asogbon et al., 2020a; Nsugbe et al., 2021a).

A number of confounding factors that impede the clinical and commercial relevance of PR-based prostheses in practical settings have been well studied with solutions proposed in recent years (Fougner et al., 2010; Lorrain et al., 2010; Tkach et al., 2010; Young et al., 2011; He et al., 2013; Qing et al., 2021). For instance, electrode shift (Young et al., 2011), muscle contraction force variation (Lorrain et al., 2010; Tkach et al., 2010), abrupt alteration in limb position (Fougner et al., 2010), and variability arising from long-term EMG recordings (He et al., 2013), etc., are confounding factors that have been researched with potential solutions proposed. Despite these advances, an essential aspect that the above factors and many others rely upon is the EMG signal recording duration (SRD) per time that may impact the characterization of finger gestures. To the best of the author's knowledge, EMG SRD has not been studied to date. For instance, when the EMG SRD is relatively lengthy, phenomena such as muscle fatigue is inevitable and may alter the muscle contraction patterns of specific limb motions; which may undermine the decoding of finger movements and, by extension, degrades the prosthesis control performance. In addition, long SRD often leads to relatively larger volume features, and classifier training time and may result in computational complexity and increased controller delay in real-time usage. On the other hand, somewhat shorter SRD may lead to reduced data collection durations that, among other advantages, allow for more convenient prosthesis recalibration protocols. On the other hand, signals acquired using short SRD may result in poor motion gesture recognition if sufficient/adequate neural information is not contained in the signal, especially if it is collected from amputee patients.

To date, several existing studies have arbitrarily employed varied myoelectric SRD for decoding targeted limb motions in the context of PR-based prostheses without taking into consideration whether or not they would yield optimal characterization of the motor tasks. For instance, Cengiz and Demir (2020) acquired myoelectric and gyroscopic signals of multi-class finger gestures with SRD of 5 s and Al-Timemy et al. (2015) investigated the influence of muscle contraction force variation on the classification performance of the EMG-PR system using SRD in the range of 8–12 s. In addition, Li et al. (2021) used an SRD of 4 s in a study aimed at enhancing the motion classification accuracy of the EMG-PR system and Samuel et al. (2018b) utilized an SRD of 5 s in a study aimed at improving the EMG-based features for prostheses control. While we acknowledge that several efforts have been made toward tackling pertinent issues in the field of PR-based prostheses technology (Cordella et al., 2016; Bates et al., 2020; Nsugbe et al., 2021b), the investigation of the optimal myoelectric SRD remains an open research question. Hence, it is essentially necessary to investigate and determine the optimal myoelectric SRD that would aid adequate characterization of amputees' limb motion intents and by extension the intuitive control of multifunctional prostheses. Also, the investigation should provide researchers and developers in the field with proper insight and guidance on the selection of appropriate SRD when conducting research or developing a requisite solution.

Therefore, this study systematically examined the impact of varied lengths of myoelectric SRD on the characterization of motor intents associated with multiple classes of fine-finger gestures performed by recruited subjects. More specifically, the experiments involved eight normally limb subjects (including six males and two females with no muscular or neurological disorder history), and each subject elicited fifteen classes (single and combined classes) of finger gestures under varying durations of EMG-SRD including 1, 5, 10, 15, and 20 s. Afterward, each SRD was pre-processed, notable feature extraction methods were applied for feature vector construction, and the feature vector was employed to build three distinct machine learning classification algorithms for the decoding of the finger gestures based on Within-SRD and Between-SRD strategies, which are described in the methodology section. Benchmark performance metrics were applied to evaluate the gesture pattern characterization and their corresponding decoding performance for each SRD.

2. Materials and methods

2.1. Data collection and processing

The myoelectric dataset utilized in this study was acquired from an online EMG datasets repository (OneDrive). The signal was collected using BagnoliTM EMG Acquisition System (manufactured by Delsys Inc.). The equipment setup and electrode placement scheme is shown in **Figures 1A, B**. Prior to the data collection process, a total of eight normally limb subjects including six males and two females with no history of muscular or neurological disorders were recruited and informed about the study's objectives (Khushaba and Kodagoda, 2012). Before their inclusion in the experiment, written informed consent was obtained from each subject, indicating their willingness to participate in the study. Afterward, eight EMG signal sensors were placed over the forearm muscles of each subject and a dual-slot adhesive skin interface was applied to firmly fix the electrodes to the skin to prevent undesirable displacement that may affect the quality of the signals. Besides, a reference electrode was placed on the wrist of each of the participants as shown in **Figure 1B**.

For the data recording task, the participants were instructed to sit down on a chair in a comfortable manner with their arms supported and fixed at a specific position (to ensure consistent arm position throughout the experiment). And fifteen classes of finger gestures were elicited as shown in **Table 1** where each motion class lasted for a period of 20 s and was followed by a rest period of 5 s. Meanwhile, EMG recordings of three experimental trials were utilized in this study (Khushaba and Kodagoda, 2012).

The recorded EMG signals were amplified using a Delsys Bagnoli-8 amplifier to a total gain of 1000 while the signal was sampled at the rate of 4000 Hz. A bandpass filter between 20 and 450 Hz and a

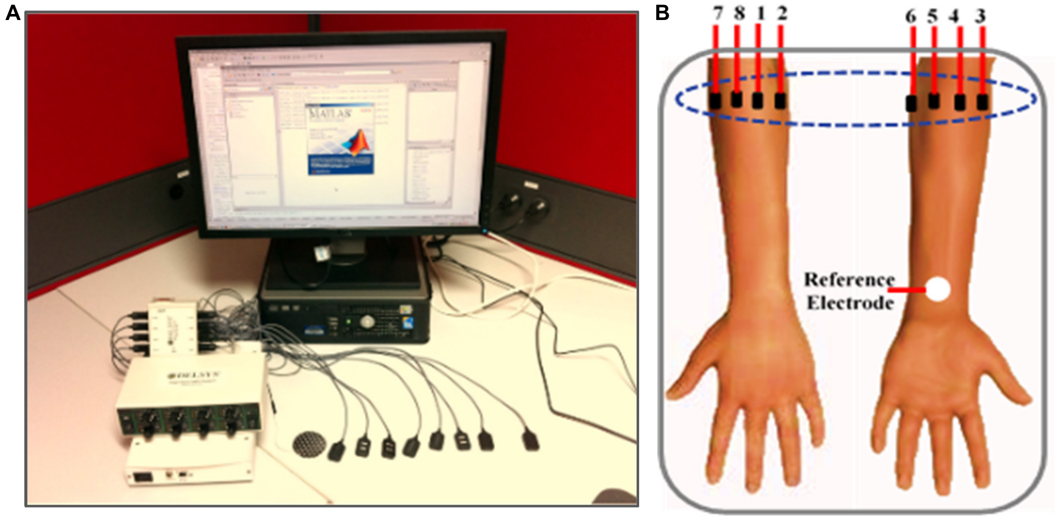


FIGURE 1
(A) EMG data acquisition system (Delsys Inc.) setup; **(B)** electrodes placements on the anterior and posterior on the participant's right arm (Khushaba and Kodagoda, 2012).

notch filter was applied to the signal to process and eliminate power line interference.

2.2. Feature extraction

To investigate the impact of SRD on the characterization of the multiple classes of figure motions, different lengths of EMG signal recordings (1, 5, 10, 15, and 20 s), as conceptualized in **Figure 2** were examined. Each SRD data was analyzed by partitioning each motion duration into a series of analysis windows with a length of 150 and 100 ms increments *via* an overlapping segmentation scheme, which has been commonly applied in the field of EMG signals processing (Englehart and Hudgins, 2003; Menon et al., 2011; Asogbon et al., 2020b). The segmentation process is often carried out to enhance the performance and response time of the PR-based myoelectric control scheme in practical settings (Asogbon et al., 2020b).

TABLE 1 Fifteen classes of finger motions with their respective codes.

Motion group	SN	Motion classes	Code
Flexion of each individual fingers	1	Thumb	T
	2	Index	I
	3	Middle	M
	4	Ring	R
	5	Little	L
Combined fingers motions	6	Thumb-index	TI
	7	Thumb-middle	TM
	8	Thumb-ring	TR
	9	Thumb-little	TL
	10	Hand close	HC
	11	Index-middle	IM
	12	Middle-ring	MR
	13	Ring-little	RL
	14	Index-middle-ring	IMR
	15	Middle-ring-little	MRL

From each analysis window segment of the EMG signal, three different features whose mathematical expressions are presented in (Eqs 1–3) were extracted individually to build a machine learning classifier for decoding the different classes of finger motions. It should be noted that the feature extraction methods have been widely applied for characterizing multiple classes of targeted limb motions and they include the Hudgins' time-domain feature set (mean absolute value: MAV, number of zero crossings: ZC, waveform length: WL, and number of slope sign changes: SSC), Novel Time-Domain Feature (NTDF, proposed by our research team), and the Root Mean Square (RMS) (Hudgins et al., 1993; Englehart and Hudgins, 2003; Samuel et al., 2018a; Asogbon et al., 2020a,b).

$$MAV = \frac{1}{k}\sum_{n=1}^{k}|x_n|$$

$$WL = \sum_{n=1}^{k-1}[(|x_{n+1} - x_n|)] \tag{1}$$

$$ZC = \sum_{n=2}^{k-1}[(x_n - x_{n-1}) * (x_n - x_{n+1})]$$

$$SSC = \sum_{n=1}^{k-1}\left[\mathrm{sgn}\,(x_n * x_{n+1}) \cap (x_n - x_{n+1}) \geq \mathrm{Thr.}\right]$$

Where $x_n$is the n^{th} sample in a given segment of the EMG recordings of length *k*. Briefly, the *MAV* represents an estimate of the mean absolute value of *x* in a given segment of the signal which is of length *k, WL* provides information regarding the wavelength characteristics in a given segment of the signal, *ZC* represents a frequency measure of the number of times the waveform crosses

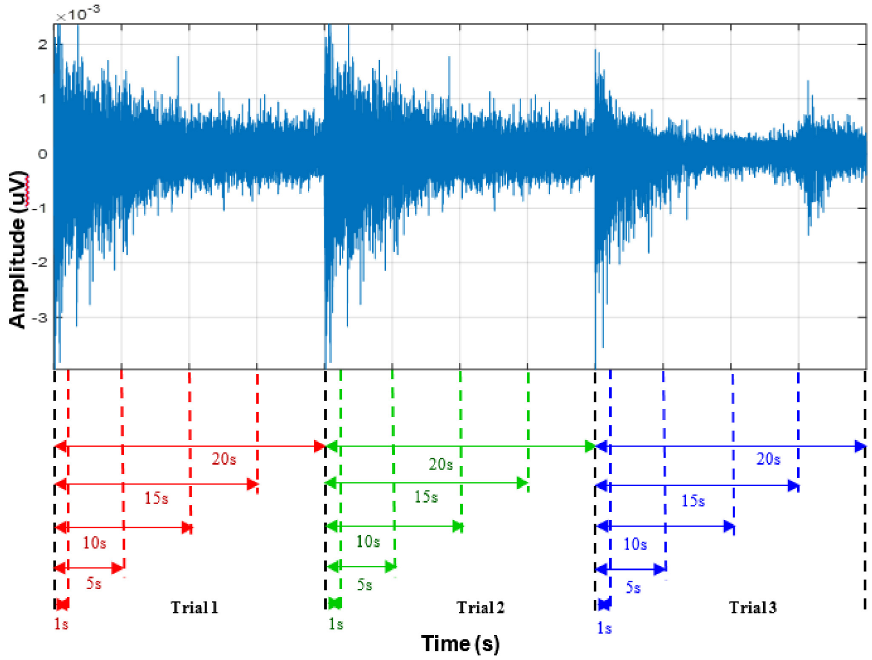


FIGURE 2
Conceptualized diagram of varying signal recording lengths considered in the study for three trials.

zero (baseline), and *SSC* denotes an alternative but complementary measure of the number of times the slope changes sign (Hudgins et al., 1993; Englehart and Hudgins, 2003). And an aggregation of these descriptors forms the TD4 feature set that was adopted in the subsequent section of the manuscript (Hudgins et al., 1993; Englehart and Hudgins, 2003). Meanwhile, the *Thr. (with a value of 0.01)* represent the threshold upon which the *SCC* value is computed.

$$SIS = \sum_{n=0}^{N-1} x[n]^2$$

$$normRSD1 = \frac{1}{N}\sum_{n=0}^{N-1} dx_1[n]^2$$

$$normRSD2 = \frac{1}{N}\sum_{n=0}^{N-1} dx_2[n]^2$$

$$normLogDet. = norm(e^{\frac{1}{N}\sum_{n=0}^{N-1}\log(x[n])}) \qquad (2)$$

$$mMSR = \frac{1}{\mathrm{k}}\sum_{\mathrm{n=1}}^{\mathrm{k}} (x_n)^{1/2}$$

$$mASM = \left| \frac{\sum_{\mathrm{n=1}}^{\mathrm{k}} (x_n)^{exp}}{k} \right|$$

$$exp = \begin{cases} 0.50, & if\ \left(n \geq 0.25 * k\ \ \&\&\ \ n \leq 0.75\right) \\ 0.75, & otherwise \end{cases}$$

Where *SIS* the denotes the simple integral square which captures the energy content in a segment of EMG signal ($x[n]/x_n$) and N denote the total length of the signal in a segment, the *normRSD*1 and *normRSD*2 represent the normalized form of the first and second order of the root squared descriptors, which captures the spectral information in a given EMG signal segment, and the *mMSR* and *mASM* descriptors capture an estimate of the power of the signal per segment (Asogbon et al., 2020a). And an aggregation of these descriptors forms the NTDF feature set that was employed subsequently (Asogbon et al., 2020a).

$$RMS = \sqrt{\frac{1}{\mathrm{k}}\sum_{\mathrm{n=1}}^{\mathrm{k}} \mathrm{x_n}^2} \qquad (3)$$

Where *RMS* denote the square root of the average power of EMG recordings (x_n) in a given segment of the signal whose length is denoted by *k*.

After each of the above-mentioned features has been extracted, three widely utilized machine learning classification algorithms with simple and intuitive structure, high accuracy, and fast computation characteristics including the Linear Discriminant Analysis (LDA), K-Nearest Neighbor (KNN), and Random Forest classifiers (Boughorbel et al., 2017; Asogbon et al., 2021) were applied to classify the motion classes for the considered SRD groups. Thus, we examined the impact of varied SRD on finger movement pattern characterization using a fixed set of features and machine learning classifiers using two approaches described below.

2.3. Data analysis and performance evaluation

The effect of SRD on EMG-PR motion intent decoding was systematically investigated based on two varied types of training and testing strategies, namely Within-SRD Group (which is represents the commonly adopted approach) and Between-SRD Approach described as follows:

(a). Within-SRD Group Scenario: In this approach, the EMG-PR scheme's performance was investigated when the training and testing feature vectors were constructed from EMG recordings of the same SRD group. Specifically, in this approach, the requisite feature vector is constructed from the first two trials (Trial 1 + Trial 2) of the recordings (designated as the training set) while the corresponding test set feature vector is obtained from the third trial (Trial 3).

(b). Between-SRD Group Scenario: In this approach, the EMG-PR scheme's performance was examined when the feature vector constructed from a specific EMG SRD group (say 1 s) is used for training the classifier while the feature vector obtained from all the SRD groups (1, 5, 10, 15, and 20 s) is used for testing the classifier's decoding performance. In addition, it should be noted that the training set is constructed from all the trials while the test set is also obtained based on all the trials.

For each of the approaches described above, evaluation metrics including classification accuracy (CA) and Mathew Correlation Coefficient (MCC) were considered and their descriptions are given as follows. The CA, a commonly used evaluation metric that represents the number of correctly classified samples over the sum of all samples [Eq. (4)] was utilized. The MCC metric which has been widely applied in multiclass problems was also adopted for evaluation in the study [Eqs (5, 6)]. MCC is considered to be a highly informative metric for assessing classification tasks since it is considered to be a balanced ratio amongst the four confusion matrix parameters (false positives, true positives, true negatives, and false negatives) (Liarokapis et al., 2014; Asogbon et al., 2020a).

$$CA = \frac{Number\ of\ correctly\ classified\ samples}{Total\ number\ of\ testing\ samples} * 100\% \tag{4}$$

$$\mathrm{MCC}_j = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{5}$$

$$\mathrm{MCC}_{ave} = \frac{\sum_{j=1}^{n} \mathrm{MCC}_j}{nclass} \tag{6}$$

Where j = 1,2,3. . . . . .number of classes (n, nclass), TP is the count of true positives, TN represents the count of true negatives, FP is the number of false positives, and FN is the number of false negatives as obtained from a confusion matrix. Meanwhile, the MCC value was computed using the macro-averaging technique.

Furthermore, the statistical significance test between the SRD groups was performed using the Friedman test with a confidence level set to $p < 0.05$. The statistical analysis was carried out in MATLAB.

3. Results

3.1. Within-SRD group scenario

3.1.1. Performance evaluation of varying signal length on EMG-PR classifier across finger gesture tasks

Utilizing different classifiers and feature sets, Figure 3 presents the average classification accuracies across the fifteen-finger motion tasks across eight subjects. The aim here is to examine the impact of individual SRD (1, 5, 10, 15, and 20 s) on the classification performance EMG-PR system. It can be observed from the result presented in Figure 3 that the average classification accuracy varies across the features and classifiers. For instance, 5 s SRD achieved the highest CAs across subjects and finger gestures with an increment ranging from 0.01 to 3.86% for the LDA classifier when compared with the performance of the other SRD groups. Similarly, some worth different phenomenon was observed when KNN and RF classifiers were employed. The 10 s (for RF) and 20 s (for KNN) outperformed other SRD groups with an increment in the classification accuracies ranging from 0.22 to 8.26% and 0.33 to 7.99%, though with an insignificant difference with the other SRD groups except for 1 s SRD. In terms of feature performance, the NTDF achieved the highest average accuracies while RMS has the lowest accuracies for all the classifiers.

Overall, the combination of LDA + NTDF recorded the highest average classification accuracy rate compared to KNN, RF, and other features. Specifically, the 1 s SRD can be observed to be the least accuracy while the other SRDs achieved higher but similar accuracies compared to 1 s SRD for the LDA-NTDF combination. Analyzing the effect of the signal recording length based on classification accuracies reported in Figure 4 for all the classifiers and features, it can be seen that the classification accuracy of 1 s SRD is significantly ($p > 0.05$) lower than the other SRDs while 5 s has the highest accuracy though with almost the same decoding performance with 10, 15 and 20 s SRD ($p > 0.05$). It is worth noting that in this study, an increase in SRD incurs increased training and testing time, which may introduce some sort of delay in the performance of the prostheses in practical deployment. In other words, the longer the recording duration the higher the computational cost and vice-versa. One possible reason for the poor performance of the 1 s SRD could be due to a lack of adequate neural information in the signal length. This result indicates that an appropriate selection of crucial parameters/methods (such as signal recording length and feature-classifier combination) would greatly impact the overall performance of the EMG-PR-based motion intent decoding strategy employed in the control of multifunctional prostheses. Across all the classifiers, statistical analyses *via* Friedman's test show no significant difference in decoding accuracies for the SRD groups for NTDF: $p = 0.11$ and TD4: $p = 0.10$, though with substantial increment between 1 s and the other SRDs. Meanwhile, there is a significant difference for RMS: $p = 0.043$.

3.2. Between-SRD group scenario

3.2.1. Effect of signal recording duration on EMG-PR classification performance

The Within-SRD Scenario has been used in many existing works; however, it may be difficult to utilize this approach to select the optimal SRD for motion intent characterization because it is not practicable in real-life situations. Hence to determine the optimal SRD, we employed the Between-SRD Approach to systematically investigate the generalizability of each SRD group for movement intent decoding. In this scenario, the machine learning classifiers were trained with data concatenated across all the trials for a specific SRD (say, 1 s) and tested using data from all trials of all the SRD groups (1, 5, 10, 15, and 20 s) and the obtained results is shown in Figures 4A–C. This Figure depicts the average CA across the 15 classes of finger gesture and participants, and it could be seen that the CA decreased for all SRDs, features, and

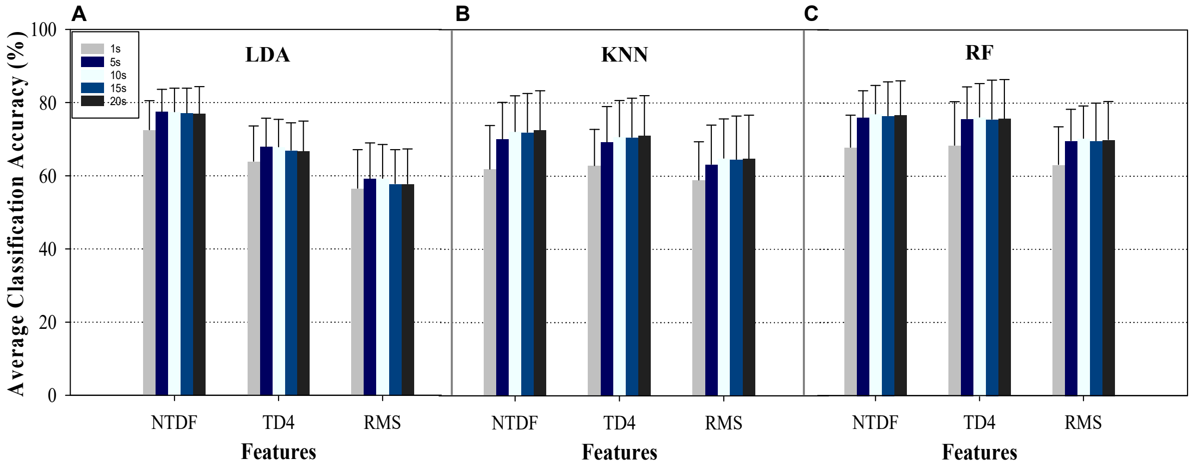


FIGURE 3
Average classification accuracies of the different groups of signal recording length based on the Within-SRD Group Scenario using **(A)** LDA, **(B)** KNN, and **(C)** RF for NTDF, TD4, and RMS features across finger gestures and participants.

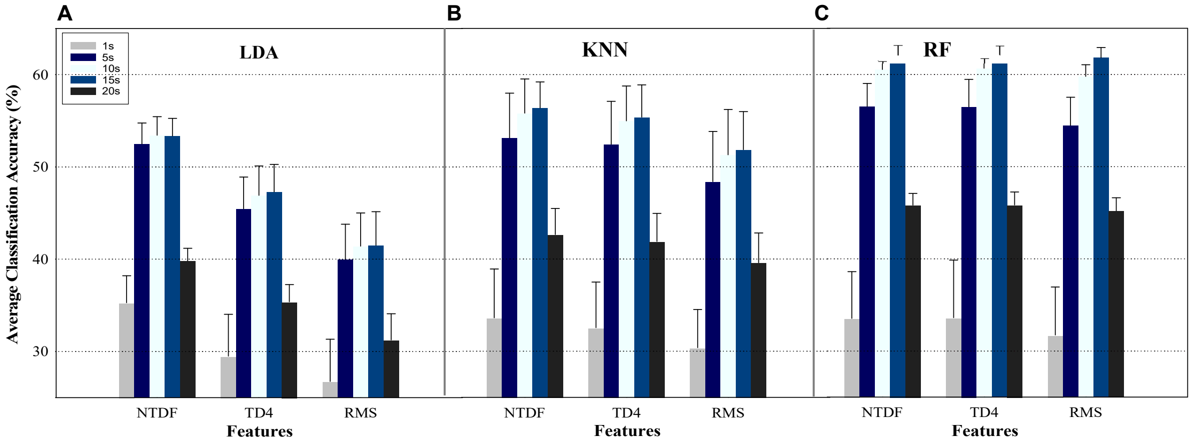


FIGURE 4
Average classification performance across finger gestures and participants based on Between-SRD Approach using the NTDF, TD4, and RMS for **(A)** LDA, **(B)** KNN, and **(C)** RF when each of the SRD was used as a training data and tested with all SRD dataset.

classifiers compared to the result presented for within-SRD scenario in Figure 3 (where the training and testing data are from the same SRD). Besides, for the Between-SRD scenario, the highest decoding performance was achieved by 15 s, followed by 10 s SRD (though with insignificant difference), while 1 and 20 s yielded the lowest CA for all features-classifier combinations. Examining the classifiers performance further, RF (Figure 4C) outperformed LDA (Figure 4A) and KNN (Figure 4B) for all the SRDs and features except for 1 s SRD for the combination of LDA + NTDF. For instance, for the most performing feature (NTDF) across motion classes and participants, RF achieved an increment of up to 7.09 and 4.71% compared to LDA and KNN, respectively, for 10 s SRD. In a similar manner, RF achieve an increment of 8.82 and 5.79% compared to LDA and KNN, correspondingly for 15 s SRD. Besides, across classifiers, the NTDF feature outperformed the other features for all the SRDs. Amid the classifiers, there are statistical significances among the SRDs for NTDF ($p = 0.021$), TD4 ($p = 0.017$), and RMS ($p = 0.017$). Similarly, significant differences occurred among the SRDs across features. From the statistical significance result, 10 and 15 s SRD achieved similar performance with no substantial difference when compared with each other. Furthermore, performance comparison between the results reported in Figure 4 for each SRD reveals that the Between-SRD Approach would significantly influence the decoding performance of the EMG-PR system.

It is worth mentioning that observations during the EMG-PR scheme's implementation revealed that the computational cost generally increase with an increase in SRD, and this may necessitate us to consider an SRD that is greater than 5 s but less than or equal to 10 s (>5 s and < = 10 s).

3.2.2. Evaluating signal duration effect based on MCC metric

The effect of the different groups of SRD on the characterization of the motor intent for the Between-SRD strategy was further examined using the MCC metric defined in section "2.3. Data analysis and performance evaluation" of the paper. Notably, the MCC is a highly informative evaluation method for estimating classification tasks mainly due to its ability to balance the ratio among the four confusion matrix parameters effectively. Thus, the corresponding MCC values for each group of signal length were computed from their respective confusion matrices.

Using the same number of classifiers and features, the obtained average MCC value across all the finger motion tasks and subjects is shown in **Figures 5A–D** using horizontal dot plots. Inspecting each of the classifier-feature combinations (**Figures 5A–C**) for all the SRD groups, NTDF yielded the highest MCC values for LDA and similar values with TD4 for KNN and RF. Meanwhile, the RMS feature has the lowest MCC values for the different groups of signal recording lengths investigated.

Observing the performance of the SRDs from **Figures 5A–C**, an overlap of symbols could be seen for 5, 10, and 15 s for LDA, and 10 and 15 s symbols overlap for KNN and RF, respectively, indicating similar MCC values. For all the classifiers and features, 1 s achieved the lowest MCC value, followed by 20 s, while 10 and 15 s SRD obtained the highest MCC values, indicating consistency with the CA results described in section "3.2.1. Decoding performance for between-SRD scenario." Compared to other classifiers, RF recorded the best values for RMS, TD4, and NTDF features. A similar performance trend of the classifiers and features for the SRD groups could be observed in the result presented in **Figures 4A–C** (see section "3.2.1. Decoding performance for between-SRD scenario"). And **Figure 5D** depicts classifier-wise computation of the mean MCC values across all features, motion tasks, and subjects. It can be seen that the RF classifier achieved the best MCC values of 0.36, 0.60, 0.65, 0.67, and 0.52 for 1, 5, 10, 15, and 20 s signal recording lengths, respectively. Furthermore, statistical analyses *via* Friedman's test show a significant difference ($p = 0.007$) in decoding accuracies across SRDs and features for LDA, KNN, and RF classifiers. Also, statistical significance occurs between the SRD groups (0.017) across classifiers.

3.2.3. Effect of signal recording length on individual finger gesture decoding

In this section, the recognition rate of each class of finger motion for individual signal recording length was examined across subjects using the combination of the NTDF feature and the RF classifier based on their performance in the Between-SRD. Utilizing line and scatter plots with error bars, the obtained result is shown in **Figures 6A–F**. It should be noted that the standard deviation across the subjects is shown with error bars in the plot. From the results, it can be seen that there are variations in the error bars for all finger gestures across the SRDs (1, 5, 10, 15, and 20 s). Specifically, in all the SRD groups, the thumb-index class (denoted by TI) recorded the highest decoding accuracies of 37.44, 60.51, 69.28, 72.36, and 60.68% for 1, 5, 10, 15, and 20 s, respectively. Also, for all the SRDs, this gesture (TI) has the highest standard error compared to other finger gestures, signifying performance variation across the participants.

On the other hand, the thumb-little (TL) finger gesture recorded the least performance for 1 s, class ring (I) for 5 s, hand close (HC) for 10 s, and thumb middle (TM) for 15 and 20 s SRD.

Comprehensively, the performance comparison between the groups revealed that the 15 s SRD group achieved the highest average recognition followed by the 10 s SRD group across all the classes of finger gestures compared to other SRDs. Meanwhile, for a clear comparison of the characterization of the motor intent across the groups, **Figure 6F** depicts the results obtained for all the groups of signal recording length. Careful analyses revealed that most classes achieved the best performance at 10 and 15 s signal recording length with most classes' symbols overlapping with one another. This result reveals that some finger gestures' decoding performance may depend on relatively longer SRDs while 1 s SRD can be seen to achieve the least decoding performance followed by 20 s SRD.

4. Discussion

Pattern recognition (PR)-based myoelectric system has been widely studied primarily because of its capability to provide control schemes that could aid seamless realization of multiple degrees of freedom functions in upper limb prosthetic technology (Cordella et al., 2016; Kuiken et al., 2016; Samuel et al., 2019; Mereu et al., 2021). In an ideal PR-based scheme, it is anticipated that repeatable muscle contraction patterns should be generated for specific limb motion tasks across trials from which feature vector of requisite motor intent is constructed and applied to build machine learning algorithms that decipher the motion intentions of amputees (Li et al., 2010; Asogbon et al., 2020a; Nsugbe et al., 2021a). Besides, various factors could affect the repeatability of muscle activation patterns even for the same limb motion task, which may dampen proper characterization of motion intent and its decoding. One of such factor that has rarely been investigated to date is the EMG-SRD employed to build the machine learning classifier meant to decode the motion task. In an attempt to gain proper insight into how EMG-SRD dynamically impacts the characterization of multiple patterns of elicited limb motion tasks, this study systematically investigated different groups of SRD (1, 5, 10, 15, and 20 s) based on the Within-SRD and Between-SRD strategies using the dataset of eight able-bodied subjects who performed fifteen classes of simple and combined finger gestures. The outcome of the investigation showed that decoding performance across subjects and finger gestures would vary depending on the EMG SRD employed regardless of the feature extraction methods and machine learning classification algorithms utilized.

The Within-SRD has been one of the popularly utilized methods for gesture recognition/classification in myoelectric- PR based systems. Through this method, several studies have reported satisfactory or high classification accuracy for either forearm or finger motion tasks using an average of ≤ 6 s SRD. Unfortunately, the high accuracies reported in all these studies have not translated into robust or intuitive prostheses control schemes that could be widely adopted in clinical and commercial settings. One possible reason may be that the commonly adopted Within-SRD approach in the existing works may not reflect what is obtainable as it relates to practical deployment of the prostheses, which led us to investigate the Between-SRD scenario in the current study. Comparing the performance of the Within-SRD (**Figures 3A–C**) and Between-SRD (**Figures 4A–C**), it is obvious that the selection of optimal EMG-SRD should be based on using a more realistic/practical approach (such as the Between-SRD) that will accommodate changes in real-life situations rather than the Within-SRD scenario that has been widely adopted.

Specifically, in the Within-SRD scenario (**Figure 3**), the average classification accuracies for motion intent decoding were mostly observed to increase with a corresponding increase in EMG signal length for all the features except for a few instances, where the performance either remains approximately the same or declines. Besides, this phenomenon was less obvious when the same set of features were employed on KNN and RF classifiers except for the NTDF feature. Unlike the RMS and TD4 features, the NTDF

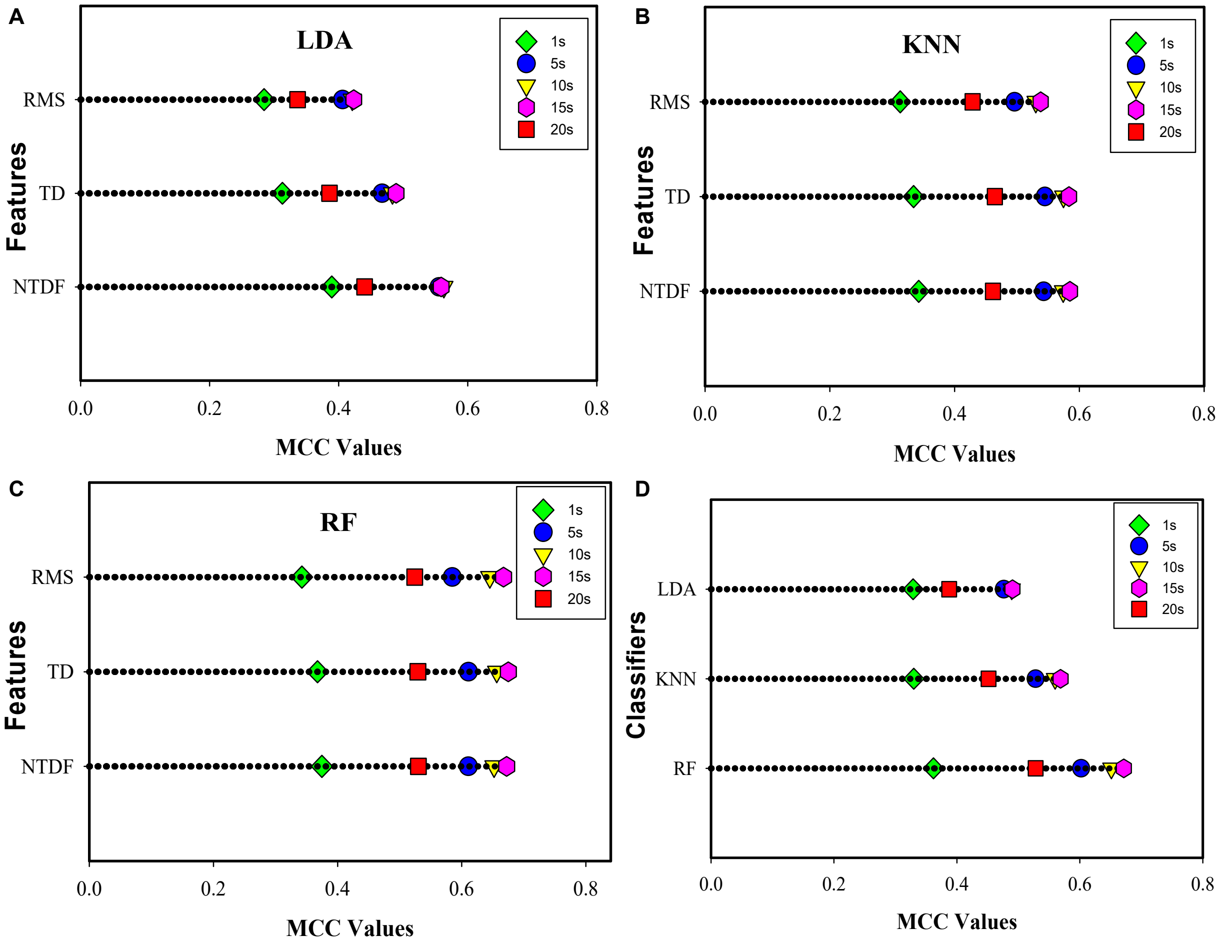


FIGURE 5
Average MCC values of the motion duration groups across finger gestures and participants based on Between-SRD Strategy using NTDF, TD4, and RMS for **(A)** LDA, **(B)** KNN, and **(C)** RF, **(D)** average MCC value using all the features, for LDA, KNN, and RF classifiers.

feature recorded relatively higher average classification accuracies especially for the LDA and KNN classifiers, demonstrating its consistency and stability capabilities. Examining the Between-SRD scenario results (Figures 4A–C), a relatively similar performance trend is seeable with the Within-SRD approach for the features (NTDF, TD4, and RMS) and Classifiers (RF, KNN, LDA) except where LDA + NTDF achieved better performance than KNN + NTDF and RF + NTDF (Figures 3A–C). However, for these two scenarios, this trend is different for the SRD groups suggesting that EMG-SRD will influence the control performance of the prosthesis system in practical applications. In Figures 4A–C, the CA increased as the EMG SRD increased and dropped significantly after 15 s for all the features and classifiers. One potential explanation for the poor performance of 1 s SRD could be that the SRD is too short, and the contained motor information in this short signal recording cannot adequately provide motor information for characterization of the finger gestures. And performance degradation in 20 s SRD could result from fatigue or lack of generalizability of relatively lengthy EMG signal recordings.

In addition to the classification accuracy metric, we investigated the impact of signal recording length on the characterization of the finger motions using the MCC metric and found that the different groups of signal length would result in the varied characterization of the corresponding classes of finger motions (Figures 5A–D). Precisely similar to the result in Figures 4A–C for the Between-SRD scenario, the MCC values increase with a corresponding increase from 1 to 15 s SRD and declined for 20 s SRD. From the plots in Figure 5, overlaps were noticeable for the MCC values of 10 and 15 s SRD, that achieved the best MCC values. Meanwhile, the 1 and 20 s SRD recorded the lowest values. And these results further substantiate the Between-SRD decoding accuracies presented in Figures 4A–C.

To understand how the SRD would impact the characterization of the individual class of finger gesture based on the Between-SRD strategy, we analyzed the fifteen classes of gestures performed across subjects for each group of SRD and observed that signal length would differently influence the classification of the gestures (Figure 6). For all the SRD groups, the thumb-index (TI) finger gesture has the highest accuracy, though with high error bars. From the further examination, the high error bars was because of high-performance variation among the subjects. For instance, participants 5, 6, and 7 recorded higher accuracies for all the SRDs except for 1 and 5 s SRD where participant 5's performance is like the others. Overall, the decoding accuracies for most finger gestures were higher for 10 and 15 s SRD, with most classes CA overlapping each other. Again, the performance of the SRD groups corroborates our earlier analyses.

Generally, for all the metrics considered in this study for the Between-SRD scenario, the 15 s SRD achieved the highest performance, followed by 10 s for the combination of RF + NTDF.

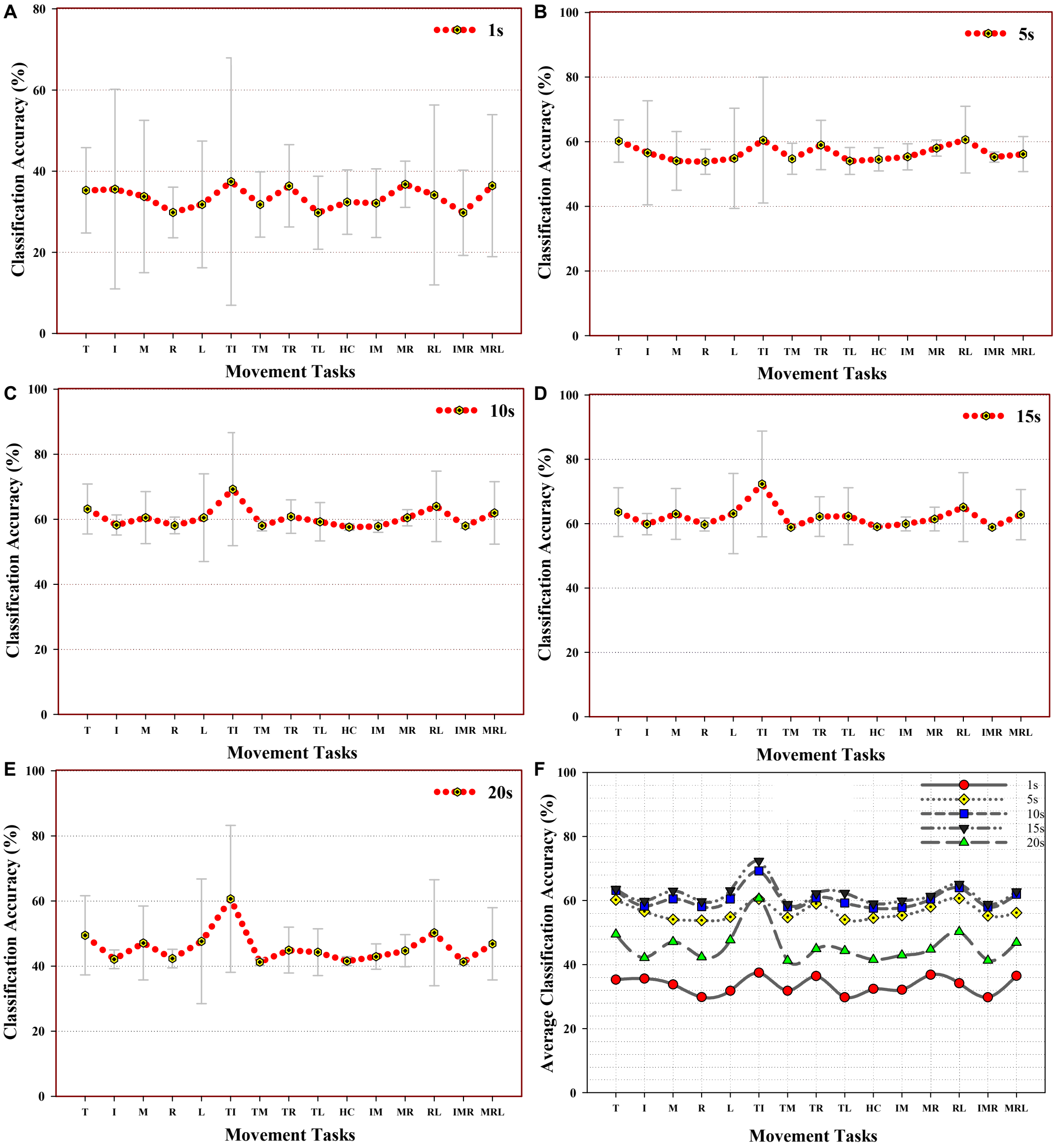


FIGURE 6
Average decoding performance across subjects for individual finger gesture using the combination of NTDF feature and LDA classifier for **(A)** 1 s, **(B)** 5 s, **(C)** 10 s, **(D)** 15 s, and **(E)** 20 s, **(F)** multiple splines curve lines and scatter plot showing performance comparison among the motion classes.

Nevertheless, it is noteworthy to state that the performance difference in these two SRD groups (10 and 15 s) is not statistically significant ($p > 0.05$), almost the same or slightly different (<1–2% for CA and MCC metrics), and also increased SRDs would lead to increase computational cost. Therefore, we would suggest a signal recording length of greater than 5 s but less than or equal to 10 s (>5 and ≤ 10 s) as being potential considering the fact that the lengthier the EMG recordings the more processing time it may require to build the classifier. Findings from our study suggest that the may be a safe zone in terms of the SRD.

Finally, the main strength of this study is that it provided a proper insight into the impact of SRD on EMG classification accuracy using three different time domain features and machine learning classification algorithms and how to select the optimal SRD that would be robust in practical situations. To the authors' knowledge, this has not been previously investigated. Moreover, it is essential to mention the drawbacks of this study. Firstly, the dataset was acquired from only healthy subjects with three trials for each class of finger gestures. Secondly, the presented experimental results were based on offline analysis. In our future work, we hope to recruit more healthy subjects and amputees from which a wider range of gestures (including finger, forearm, and wrist movements) will be obtained to further validate our hypothesis. It also worth mentioning that our future investigations shall be done in an online setting other than the offline analysis carried out in the current study, employing real-time evaluation metrics to further validate our hypothesis.

5. Conclusion

Pattern recognition-based electromyogram control method for prostheses has been highlighted and demonstrated as a potential control strategy that can aid the realization of multiple degrees of freedom prosthetic functions in a dexterous manner. Besides, an important aspect of the framework that has rarely been investigated is the myoelectric SRD upon which multiple classes of limb motion tasks are characterized. Thus, this study systematically investigated the impact of varying EMG-SRD on the characterization of motor intents associated with multiple classes of finger gestures. The investigation involved characterizing fifteen classes of finger gestures performed by eight normally limb subjects under varying lengths of EMG-SRD (1, 5, 10, 15, and 20 s). Thereafter, each group of recordings was pre-processed followed by the extraction of different feature sets, and machine learning classification algorithms were employed for decoding the corresponding gestures based on two strategies namely Between-SRD and Within-SRD scenarios. Comparison between these scenarios revealed that EMG-SRD would influence the performance of motion intent decoding. In the experimental results for Between-SRD scenario, SRDs of 10 and 15 s yielded reasonably decent performance compared to other SRDs in terms of movement intent decoding across finger gestures, subjects, feature sets, and classifiers. Considering the increased computation complexity that comes with increased SRD and the fact that no significant/substantial improvement was seen in the performance of 10 and 15 s SRD, the study will recommend a longer SRD (>5 and ≤ 10 s) depending on the research objective. The optimal SRD was determined based on the Between-SRD approach because it is more realistic/practicable compared to the Within-SRD scenario. More importantly, determining the optimal signal length is crucial to adequately characterize multiple classes of targeted limb motions in the context of EMG-PR-based control for multifunctional prostheses. In our future work, further investigations will be conducted to validate the findings of this study.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://www.rami-khushaba.com/biosignals-repository.

Ethics statement

The studies involving human participants were reviewed and approved by we used a public dataset whose URL is: https://www.rami-khushaba.com/biosignals-repository. The patients/participants provided their written informed consent to participate in this study.

Author contributions

MA, OS, and GL: conceptualization, funding acquisition, and writing – review and editing. MA and OS: data curation, writing – original draft, and methodology. OS and GL: supervision. YL and DM: further data analysis and result interpretation. EN, FK, and SC: validation. MA, OS, SC, EN, and FK: visualization. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China Grants (#U1913601, #62150410439, and #82050410452), Ministry of Science and Technology, Shenzhen (#QN2022032013L), Shenzhen Governmental Basic Research Grant (#JCYJ20180507182508857), and Shenzhen Governmental Collaborative Innovation Program (#SGLH20180625142402055), and National Key R&D Program of China (#2022YFE0197500).

Acknowledgments

We would like to thank Rami Khushaba and Sarath Kodagoda for the acquisition of the dataset used in this study.

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

References

Al-Timemy, A. H., Khushaba, R. N., Bugmann, G., and Escudero, J. (2015). Improving the performance against force variation of EMG controlled multifunctional upper-limb prostheses for transradial amputees. *IEEE Trans. Neural Syst. Rehabil. Eng.* 24, 650–661. doi: 10.1109/TNSRE.2015.2445634

Asogbon, M. G., Samuel, O. W., Geng, Y., Oluwagbemi, O., Ning, J., Chen, S., et al. (2020a). Towards resolving the co-existing impacts of multiple dynamic factors on the performance of EMG-pattern recognition-based prostheses. *Comput. Methods Prog. Biomed.* 184:105278. doi: 10.1016/j.cmpb.2019.105278

Asogbon, M. G., Samuel, O. W., Jiang, Y., Wang, L., Geng, Y., Sangaiah, A. K., et al. (2020b). Appropriate feature set and window parameters selection for efficient motion intent characterization towards intelligently smart EMG-PR system. *Symmetry* 12:1710.

Asogbon, M. G., Samuel, O. W., Li, X., Nsugbe, E., Scheme, E., and Li, G. (2021). A linearly extendible multi-artifact removal approach for improved upper extremity EEG-based motor imagery decoding. *J. Neural Eng.* Online ahead of print. doi: 10.1088/1741-2552/ac0a55

Bates, T. J., Fergason, J. R., and Pierrie, S. N. (2020). Technological advances in prosthesis design and rehabilitation following upper extremity limb loss. *Curr. Rev. Musculoskeletal Med.* 13, 485–493. doi: 10.1007/s12178-020-09656-6

Boughorbel, S., Jarray, F., and El-Anbari, M. (2017). Optimal classifier for imbalanced data using matthews correlation coefficient metric. *PLoS One* 12:e0177678. doi: 10.1371/journal.pone.0177678

Cengiz, T., and Demir, M. C. (2020). Detection and classification of muscle activation in EMG data acquired by Myo armband. *Avrupa Bilim ve Teknoloji Dergisi* 178–183. doi: 10.31590/ejosat.779660

Cordella, F., Ciancio, A. L., Sacchetti, R., Davalli, A., Cutti, A. G., Guglielmelli, E., et al. (2016). Literature review on needs of upper limb prosthesis users. *Front. Neurosci.* 10:209. doi: 10.3389/fnins.2016.00209

Englehart, K., and Hudgins, B. (2003). A robust, real-time control scheme for multifunction myoelectric control. *IEEE Trans. Biomed. Eng.* 50, 848–854. doi: 10.1109/TBME.2003.813539

Fougner, A., Scheme, E., Chan, A., Englehart, K., and Stavdahl, Ø. (2010). Resolving the limb position effect in myoelectric pattern recognition,. *IEEE Trans. Neural Syst. Rehabil. Eng.* 19, 644–651. doi: 10.1109/TNSRE.2011.2163529

He, J., Zhang, D., Sheng, X., and Zhu, X. (2013). "Effects of long-term myoelectric signals on pattern recognition," in *Intelligent Robotics and Applications*, (Berlin: Springer), 396–404. doi: 10.1007/978-3-642-40852-6_40

Hudgins, B., Parker, P., and Scott, R. N. (1993). A new strategy for multifunction myoelectric control. *IEEE Trans. Biomed. Eng.* 40, 82–94. doi: 10.1109/10.204774

Khushaba, R. N., and Kodagoda, S. (2012). "Electromyogram (EMG) feature reduction using mutual components analysis for multifunction prosthetic fingers control," in *2012 12th International Conference on Control Automation Robotics & Vision (ICARCV)*, Piscataway, NJ: IEEE, 1534–1539. doi: 10.1109/ICARCV.2012.6485374

Kuiken, T. A., Miller, L. A., Turner, K., and Hargrove, L. J. (2016). A comparison of pattern recognition control and direct control of a multiple degree-of-freedom transradial prosthesis. *IEEE J. Trans. Eng. Health Med.* 4, 1–8. doi: 10.1109/JTEHM.2016.2616123

Li, G., Schultz, A. E., and Kuiken, T. A. (2010). Quantifying pattern recognition—Based myoelectric control of multifunctional transradial prostheses. *IEEE Trans. Neural Syst. Rehabil. Eng.* 18, 185–192. doi: 10.1109/TNSRE.2009.2039619

Li, X., Tian, L., Zheng, Y., Samuel, O. W., Fang, P., Wang, L., et al. (2021). A new strategy based on feature filtering technique for improving the real-time control performance of myoelectric prostheses. *Biomed. Signal Process. Control* 70:102969. doi: 10.1016/j.bspc.2021.102969

Liarokapis, M., Kyriakopoulos, K. J., and Artemiadis, P. (2014). "A learning scheme for EMG based interfaces: on task specificity in motion decoding domain," in *Neuro-Robotics*, Dordrecht: Springer, 3–36. doi: 10.1007/978-94-017-8932-5_1

Lorrain, T., Jiang, N., and Farina, D. (2010). Influence of the training set on the accuracy of surface EMG classification in dynamic contractions for the control of multifunction prostheses. *J. Neuro Eng. Rehabil.* 8:25. doi: 10.1186/1743-0003-8-25

Menon, R., Di, C. G., Lakany, H., Petropoulakis, L., Conway, B. A., and Soraghan, J. J. (2011). Study on interaction between temporal and spatial information in classification of EMG signals for myoelectric prostheses. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 1832–1842. doi: 10.1109/TNSRE.2017.2687761

Mereu, F., Leone, F., Gentile, C., Cordella, F., Gruppioni, E., and Zollo, L. (2021). Control strategies and performance assessment of upper-limb TMR prostheses: a review. *Sensors* 21:1953. doi: 10.3390/s21061953

Nsugbe, E., Samuel, O. W., Asogbon, M. G., and Li, G. (2021a). Contrast of multi-resolution analysis approach to transhumeral phantom motion decoding. *CAAI Trans. Intell. Technol.* 6, 360–375. doi: 10.1049/cit2.12039

Nsugbe, E., Samuel, O. W., Asogbon, M. G., and Li, G. (2021b). Phantom motion intent decoding for transhumeral prosthesis control with fused neuromuscular and brain wave signals. *IET Cyber Syst. Robot.* 3, 77–88. doi: 10.1049/csy2.12009

OneDrive (2013). *OneDrive*. Available online at: https://onedrive.live.com/?authkey=%21Ar1wo75HiU9RrLM&id=AAA78954F15E6559%21312&cid=AAA78954F15E6559. (accessed May 12, 2022).

Parajuli, N., Sreenivasan, N., Bifulco, P., Cesarelli, M., Savino, S., Niola, V., et al. (2019). Real-time EMG based pattern recognition control for hand prostheses: a review on existing methods, challenges and future implementation. *Sensors* 19:4596. doi: 10.3390/s19204596

Qing, Z., Lu, Z., Cai, Y., and Wang, J. (2021). Elements influencing sEMG-based gesture decoding: muscle fatigue. Forearm angle and acquisition time. *Sensors* 21:7713. doi: 10.3390/s21227713

Samuel, O. W., Asogbon, M. G., Geng, Y., Al-Timemy, A. H., Pirbhulal, S., Ji, N., et al. (2019). Intelligent EMG pattern recognition control method for upper-limb multifunctional prostheses: advances, current challenges, and future prospects. *IEEE Access* 7, 10150–10165. doi: 10.1109/ACCESS.2019.2891350

Samuel, O. W., Zhou, H., Li, X., Wang, H., Zhang, H., Sangaiah, A. K., et al. (2018b). Pattern recognition of electromyography signals based on novel time domain features for amputees' limb motion classification. *Comput. Electr. Eng.* 67, 646–655. doi: 10.1016/j.compeleceng.2017.04.003

Samuel, O. W., Asogbon, M. G., Geng, Y., Chen, S., Feng, P., Chuang, L., et al. (2018a). "A novel time-domain descriptor for improved prediction of upper limb motion intent in EMG-PR system," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Piscataway, NJ: IEEE, 3513–3516. doi: 10.1109/EMBC.2018.8513015

Tkach, D., Huang, H., and Kuiken, T. A. (2010). Study of stability of time-domain features for electromyographic pattern recognition. *J. Neuroeng. Rehabil.* 7, 1–13. doi: 10.1186/1743-0003-7-21

Vujaklija, I., Farina, D., and Aszmann, O. C. (2016). New developments in prosthetic arm systems. *Orthoped. Res. Rev.* 8, 31–38. doi: 10.2147/ORR.S71468

Wheaton, L. A. (2017). "Neurorehabilitation in upper limb amputation: understanding how neurophysiological changes can affect functional rehabilitation. *J. Neuroeng. Rehabil.* 14, 1–12. doi: 10.1186/s12984-017-0256-8

Young, A. J., Hargrove, L., and Kuiken, T. (2011). The effects of electrode size and orientation on the sensitivity of myoelectric pattern recognition systems to electrode shift. *IEEE Trans. Biomed. Eng.* 58, 2537–2544. doi: 10.1109/TBME.2011.2159216

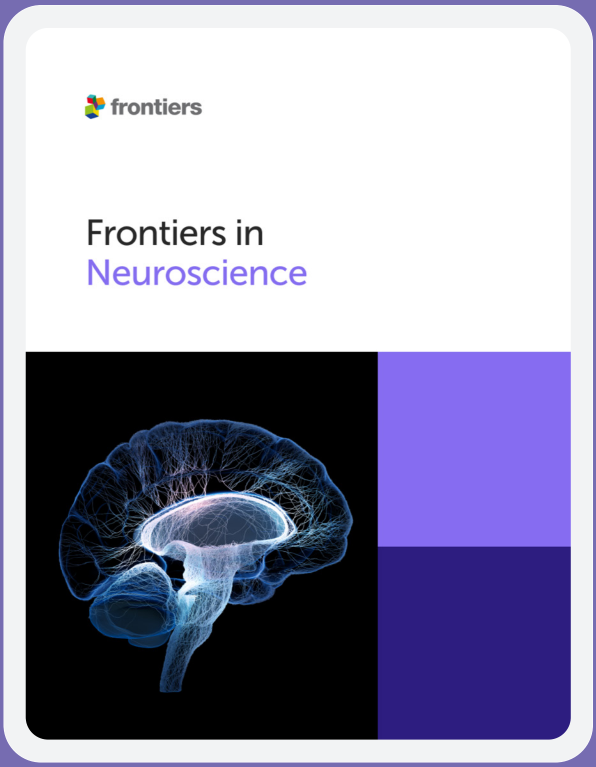
Frontiers in
Neuroscience

Provides a holistic understanding of brain function from genes to behavior

Part of the most cited neuroscience journal series which explores the brain - from the new eras of causation and anatomical neurosciences to neuroeconomics and neuroenergetics.

Discover the latest Research Topics

See more →



frontiers

Frontiers in
Neuroscience

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact