ENABLING WEARABLE BRAIN TECHNOLOGIES - METHODS AND APPLICATIONS

EDITED BY: Marcin Wozniak, Victor Hugo C. de Albuquerque and Adel Said Elmaghraby

PUBLISHED IN: Frontiers in Human Neuroscience

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88971-578-7
DOI 10.3389/978-2-88971-578-7

About Frontiers

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

Frontiers Journal Series

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

Dedication to Quality

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

What are Frontiers Research Topics?

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

ENABLING WEARABLE BRAIN TECHNOLOGIES - METHODS AND APPLICATIONS

Topic Editors:
Marcin Wozniak, Silesian University of Technology, Poland
Victor Hugo C. de Albuquerque, University of Fortaleza, Brazil
Adel Said Elmaghraby, University of Louisville, United States

Citation: Wozniak, M., de Albuquerque, V. H. C., Elmaghraby, A. S., eds. (2021). Enabling Wearable Brain Technologies - Methods and Applications. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-578-7

Table of Contents

frontiers in Human Neuroscience

EDITORIAL
published: 10 September 2021
doi: 10.3389/fnhum.2021.722388

Editorial: Enabling Wearable Brain Technologies - Methods and Applications

Marcin Wozniak[1]*, *Victor Hugo C. de Albuquerque*[2]* and *Adel Said Elmaghraby*[3]*

[1] Faculty of Applied Mathematics, Silesian University of Technology, Gliwice, Poland, [2] Graduate Program on Teleinformatics Engineering, Federal University of Ceará, Fortaleza, Brazil, [3] Computer Science and Engineering, JB Speed School of Engineering, University of Louisville, Louisville, KY, United States

Keywords: computational intelligence, computational intelligence method, brain understanding and analysis, augmented reality and human machine interaction, augmented reality, deep learning, deep learning-artificial neural network, internet of wearable brain things

Editorial on the Research Topic

Enabling Wearable Brain Technologies - Methods and Applications

INTRODUCTION

Modern technological advances in portable technologies have made it possible to study the brain in its entirety, to promote convalescence after an accident and to help in the treatment of various brain diseases and disorders, to develop a greater perception of the world around us, etc. All of this is possible because the results of modern technologies allow us to better understand how the brain works. Readings and images from different sensors can be processed as digital values (Zhu et al.; Zheng et al.; Zhang et al.), dedicated image forms or even sophisticated waves using advanced methodologies (Wang et al.; Zhang et al.). A variety of new aspects is studied in modern sciences to better understand how the brain works.

This Research Topic aims to find answers to the following questions: Is it possible to use brain to control some devices? How the processes of control and interaction through modern technology reflect thinking and functions of the brain? Do we see any clues for potential disorder from readings of brain sensors? Do we have yet good methods to analyze brain signals and use them for prediction? Is modern science able to simulate brain and its functions?

This Research Topic presents a collection of research papers covering an open cross-field junction between technology, methodology, science and didactics to enable professional discussion and presentation of innovative and efficient ideas to maximize any possible benefits of the research to the society, at a technological and methodological level.

RESEARCH TOPIC COVERAGE

This Research Topic received a wide acknowledgments among research communities in various parts of the World. The authors of accepted publications presented articles covering latest advances on Internet of Things, deep learning for neuroimage and neurosignal processing, Virtual and Augmented Reality for neuroscience, multimodal data fusion, and various interesting real world applications and simulation.

In Wang et al. was presented a research on the model of Electroencephalography (EEG) signal from electrodes placed on the scalp. Experiment has shown how to approach potential disorders of the auditory function within the brain by using proposed Random Stimulation Rate (RSR) method

OPEN ACCESS

Edited and reviewed by:
Gernot R. Müller-Putz,
Graz University of Technology, Austria

***Correspondence:**
Marcin Wozniak
marcin.wozniak@polsl.pl
Victor Hugo C. de Albuquerque
victor.albuquerque@ieee.org
Adel Said Elmaghraby
adel@ieee.org

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

Received: *08 June 2021*
Accepted: *09 August 2021*
Published: *10 September 2021*

Citation:
Wozniak M, de Albuquerque VHC and Elmaghraby AS (2021) Editorial: Enabling Wearable Brain Technologies - Methods and Applications. Front. Hum. Neurosci. 15:722388. doi: 10.3389/fnhum.2021.722388

which integrates a random interval between two adjacent stimuli. As a result an early diagnoses of auditory pathway abnormities was proposed.

In Flynn et al. was presented a model of automatic emotion recognition by using neural network approach. In the research Authors analyzed combination of methodology underpinned by psychology and latest technology like iMotions biometric research platform to formulate model of relations between automatically and interactively captured responses of participants in a form of International Affective Picture System (IAPS) standard.

In Zhu et al. was proposed Hierarchical Attentive Decoding (HAD) model where by exploring knowledge associations in sequence-to-sequence classification framework Authors proposed a model of adaptive assignment of multiple species from input texts.

In Zhang et al. was presented a new idea for automatic diagnosis of patients with alcoholism by analyzing neural activity from electroencephalogram (EEG) signals in a two-dimensional perspective. Model works in a form of Transfer Learning with Convolutional Neural Network (CNN). Authors tested various configurations to find the best architecture.

In Zheng et al. was presented a model to estimate mental fatigue by proposed Steady-State Visual Evoked Potential (SSVEP) from six stimulus paradigms: reverse vertical sinusoidal gratings, reverse horizontal sinusoidal gratings, reverse vertical square-wave gratings, brief-onset vertical sinusoidal gratings, reversal checkerboards, and oscillating expansion–contraction concentric rings.

In Ptito et al. was presented a discussion on new trends and possibilities sourced in research on technology and neuroscience to help blind people by using sensory substitution or cross-modal plasticity for brain to support vision sense.

In Zhang et al. was presented a novel non-contact sensing technique to detect sensory ataxia and cerebellar ataxia, which reveal in poor body coordination and balance disorder. Authors present Romberg's test and gait analysis model which collect information for machine learning based classifiers. Results have shown high potential for further research and development.

All presented articles show important ideas for solving medical problems in brain sensing and activity by applied models of Artificial Intelligence. Editors are happy to present this collection to the scientific community, and hope that it will start a new trend on neuroscience development.

AUTHOR CONTRIBUTIONS

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

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

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

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 20 March 2020
doi: 10.3389/fnhum.2020.00078

The Effects of Random Stimulation Rate on Measurements of Auditory Brainstem Response

Xin Wang[1,2†], Mingxing Zhu[1,2†], Oluwarotimi Williams Samuel[1], Xiaochen Wang[1,2], Haoshi Zhang[1], Junjie Yao[3], Yun Lu[4], Mingjiang Wang[4], Subhas Chandra Mukhopadhyay[5], Wanqing Wu[6], Shixiong Chen[1] and Guanglin Li[1]*

[1]CAS Key Laboratory of Human–Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences (CAS), Shenzhen, China, [2]Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen, China, [3]The Duke Institute for Brain Sciences, Duke University, Durham, NC, United States, [4]The School of Electronics and Information Engineering, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, China, [5]The Department of Engineering, Macquarie University, Sydney, NSW, Australia, [6]The School of Biomedical Engineering, Sun Yat-Sen University, Guangzhou, China

Electroencephalography (EEG) signal is an electrophysiological recording from electrodes placed on the scalp to reflect the electrical activities of the brain. Auditory brainstem response (ABR) is one type of EEG signals in response to an auditory stimulus, and it has been widely used to evaluate the potential disorders of the auditory function within the brain. Currently, the ABR measurements in the clinic usually adopt a fixed stimulation rate (FSR) technique in which the late evoked response could contaminate the ABR signals and deteriorate the waveform differentiation after averaging, thus compromising the overall auditory function assessment task. To resolve this issue, this study proposed a random stimulation rate (RSR) method by integrating a random interval between two adjacent stimuli. The results showed that the proposed RSR method was consistently repeatable and reliable in multiple trials of repeated measurements, and there was a large amplitude of successive late evoked response that would contaminate the ABR signals for conventional FSR methods. The ABR waveforms of the RSR method showed better wave I–V morphology across different stimulation rates and stimulus levels, and the improved ABR morphology played an important role in early diagnoses of auditory pathway abnormities. The correlation coefficients as functions of averaging time showed that the ABR waveform of the RSR method stabilizes significantly faster, and therefore, it could be used to speed up current ABR measurements with more reliable testing results. The study suggests that the proposed method would potentially aid the adequate reconstruction of ABR signals towards a more effective means of hearing loss screening, brain function diagnoses, and potential brain–computer interface.

Keywords: auditory brainstem response, random stimulation rate, hearing loss, hearing impairment, electroencephalogram

OPEN ACCESS

Edited by:
Victor Hugo C. de Albuquerque, University of Fortaleza, Brazil

Reviewed by:
Luca Vollero, Campus Bio-Medico University, Italy
Sandeep Pirbhulal, University of Beira Interior, Portugal

***Correspondence:**
Shixiong Chen
sx.chen@siat.ac.cn

†These authors have contributed equally to this work

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

Received: *23 December 2019*
Accepted: *21 February 2020*
Published: *20 March 2020*

Citation:
Wang X, Zhu M, Samuel OW, Wang X, Zhang H, Yao J, Lu Y, Wang M, Mukhopadhyay SC, Wu W, Chen S and Li G (2020) The Effects of Random Stimulation Rate on Measurements of Auditory Brainstem Response. Front. Hum. Neurosci. 14:78. doi: 10.3389/fnhum.2020.00078

INTRODUCTION

Hearing impairment or hearing loss primarily occurs as a result of damage to a specific part of the ear due to congenital defects, diseases, exposure to excessively loud noise, or injury, among others. This phenomenon often leads to a decrease in the auditory sensitivity or hearing dysfunction that prevents humans from sensing sounds in their environment.

Individuals suffering from hearing impairment usually have difficulties in adequately perceiving and understanding what is spoken around them. Depending on where the damage occurs, hearing loss can be classified into different types that include conductive (outer-ear or middle-ear problem), sensory (inner-ear problem), neurological (auditory center problem), and mixed hearing loss (Elzouki et al., 2012). From recent studies, it was reported that hearing loss affects over 1.1 billion individuals across different age groups (World Health Organization, 2011; Olusanya et al., 2014; Vos et al., 2015). It leads to disability in about 50% (360–538 million) of the hearing loss populace with around 124 million persons having moderate to severe disability (World Health Organization, 2015). Meanwhile, it has been projected that the number of persons with hearing loss will continually increase with time since the number of affected individuals rose from 1.1 billion in 2013 to 1.4 billion in 2017 (James et al., 2018). Due to the growing number of patients, more and more attention are being paid to medical technology research. Some researchers pay attention to the construction of a physiological system platform and the development of some acquisition methods (Samuel et al., 2017a,b; Pirbhulal et al., 2018; Wu et al., 2018; de Oliveira et al., 2019). However, researches on the system or platform cannot solve the essential problems in clinically used technology and help less with the early detection in an auditory aspect. Therefore, there is an urgent need to develop an efficient strategy for early detection and timely treatment of hearing loss, to prevent the language development impediment of newborns and life quality decline of adults.

In clinical settings, the commonly used method for auditory function assessment is the auditory brainstem response (ABR) measurement, in which electrodes are placed on the scalp to record electrical brain activities in response to brief sound stimulation to the ear. Fundamentally, the ABR signal usually occurs within the first 10 ms following the stimulus onset, and it could reflect the functional status of the auditory pathway when the neural information of the incoming sound propagates from the auditory nerve to the auditory cortex. Therefore, the ABR measurement is commonly regarded as an objective method for evaluating the perceiving sensitivity of the auditory system (Galambos and Hecox, 1978; Avan and Bonfils, 1997; Alwan, 2012; Xie et al., 2018). ABR parameters such as the amplitude of peaks, the latency of waves I–V, interpeak latency, and interaural latency are very important for the detection of brainstem impairments and central auditory abnormities. For example, the absolute latency of wave I is prolonged, but interpeak latencies are not affected for conductive hearing loss. In contrast, wave I tends to be normal, but the interpeak latencies of waves I–III and I–IV are usually prolonged for neural hearing loss.

The lowest intensity at which wave V of ABR signals can be reliably observed could provide an estimation of the hearing threshold. However, it is still controversial to use the ABR thresholds as a replacement of the behavior pure-tone audiometry (PTA). For instance, Canale et al. (2020) reported that the mean difference between the ABR and PTA thresholds was about 20 dB in normal hearing, and no differences were found in conductive or sensorineural hearing loss. Ceylan et al. (2018) showed that the mean threshold difference was 5 dB at 1 kHz and that there was no significant difference at high frequencies. Lu et al. (2017) reported that the click-evoked ABR thresholds and PTA thresholds differed by less than 20 dB in 72.6% subjects at frequencies of 2–4 kHz. Hoda et al. (2019) showed that there was a high degree of correlation between click ABR and behavioral PTA thresholds. However, Talaat et al. (2020) claimed that the click- and tone burst-evoked ABR hearing thresholds significantly overestimate the behavioral threshold.

Among all the factors that may lead to the controversies of comparison between ABR and PTA thresholds, the most important factor might be that it is still a great challenge to obtain reliable ABR waveform morphologies for accurate diagnostic purposes, given that the ABR signal is rather low in amplitude (as low as 0.1 μV). For instance, the current method that adopts an overlapping averaging technique with a fixed stimulation rate (FSR) usually generates a relatively poor ABR waveform when the stimulation rate is high. Although the irrelevant noises could be attenuated during the averaging, the obtained ABR signals by the FSR method not only contain the target signals but also include evoked potential trails from the previous segments. The auditory evoked potential (AEP) is segmented in three parts, namely, according to different latencies: short latency response (SLR; 0–10 ms), middle latency response (MLR; 10–50 ms), and long latency response (LLR; >50 ms). The first part SLR is characterized by a lower amplitude, which is filtered to eliminate the power frequency and other environmental interferences and averaged *via* the overlapping technique to obtain the actual ABR signal (Wong and Bickford, 1980; Aimoni et al., 2010; Rouillon et al., 2016; Jiang et al., 2018). It should be noted that the temporal gap between two adjacent stimuli onset could be short for high stimulation rate, so the unwanted components from the previous late evoked response (MLR and LLR) would mix with the ABR signal of the current stimulation. For the FSR method, the interferences from the previous MLR and LLR components would be enhanced after fixed-rate averaging, leading to undesired ABR morphology alteration.

To reduce the interferences from the MLR and LLR components, Alvarez used an iterative randomized stimulation and averaging (iterative-RSA) and deconvolution method to measure ABR at a high stimulation rate (Alvarez et al., 2010; Valderrama et al., 2012, 2014). Their method could help to reduce the late response interferences through an iterative process in the time domain. However, the deconvolution algorithms (such as least-squares deconvolution and continuous-loop averaging deconvolution) involve complex computations which are rather time-consuming. Moreover, the deconvolution algorithms require a controlling factor alpha, and the deconvolution algorithms might be unstable if the factor is not chosen properly. Talaat et al. (2020) utilized a chirp signal with the latencies of different frequency components adjusted according to the traveling wave delays of the basilar membrane and found that the chirp-evoked ABR could achieve statistically higher amplitudes within a shorter time. However, the traveling wave delays of the basilar membrane may be highly individual dependent, and the constructed chirp stimulus was usually

much longer than the commonly used click sounds, making the maximum stimulation rate largely limited by the stimulus duration. Hence, it is necessary to develop a practical method that could effectively eliminate the influences of MLR and LLR components to obtain reliable ABR waveform morphologies.

The purpose of this study is to propose a random stimulation rate (RSR) technique with the capability of mitigating the interferences arising from the late evoked response from previous stimulation. The performance of the proposed method in improving the ABR waveform morphologies and signal quality was thoroughly evaluated and compared with that of the commonly used FSR method under various stimulation conditions. It is believed that the outcome of this study may potentially aid adequate improvement of ABR measurements towards more accurate hearing loss assessment.

MATERIALS AND METHODS

Subjects

A total of 10 healthy subjects with ages from 20 to 30 years old were recruited in this study (mean age = 24 ± 2.87). The subjects had no history of outer- or middle-ear problems and had normal hearing function with thresholds of 20 dB hearing level or less for frequencies from 250 to 8,000 Hz in standard audiogram tests. The experimental sessions were conducted in an electromagnetically shielded room to prevent acoustical and electromagnetic interferences. The subjects were instructed to sit in a comfortable position and to be as quiet as possible during the test to minimize artifact interferences. The protocol of this study was approved by the Institutional Review Board of Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences (SIAT-IRB-190615-H0352).

Experimental Principles

In this study, the method of RSR was proposed to improve the signal quality of ABR measurements. To prevent the ABR signals from being contaminated by interferences from MLR and LLR components of the previous stimulation, the RSR method introduced random time intervals between two adjacent stimulations, as compared with the commonly used FSR method with a fixed stimulus onset interval (**Figure 1**). As shown in **Figure 1**, the responses of each stimulation would be averaged in reference to the stimulus onset to obtain the ABR signals, for both the FSR and RSR methods. In **Figure 1A**, the ABR evoked by the second stimulation overlapped with the late response (MLR and LLR) of the first stimulation, as indicated by the shadow area of the late response. For the FSR method, the overlapped shadowed interferences would be in synchronization with the stimulus onset and therefore be enhanced during the averaging, leading to undesired ABR waveform changes after mixing with the early ABR. In contrast, the shadowed interferences of the RSR method (**Figure 1B**) no longer synchronized with the stimulus onset after random intervals were introduced, making the shadowed interferences cancel out each other after the averaging. In this way, the interferences of the late responses could be prevented to obtain more accurate ABR results for the proposed RSR method. In **Figure 1B**, the random interval was set to duration with a uniform distribution between 0 and 10 ms. The stimulation period T was changed from 20 to 70 Hz to systemically investigate the performance of the RSR method under different conditions.

Different stimuli such as clicks, chirps, and tone bursts could be used to evoke ABR signals. The click stimulus is considered the most efficient stimulus for the ABR test due to its easy generation, short duration, and broadband (Eggermont and Moore, 2012; Lu et al., 2017). In this study, the click-based stimulus was adopted for conducting all the ABR tests. The duration of the click stimulus was set to 100 μs throughout the experiments. The earphones that were used to play the click sound were ER-2A (Etymotic Research, Inc., Elk Grove Village, IL, USA), in which plastic tubes were used to connect the earphones and the inserted earplugs to minimize the electromagnetic interferences picked up by the electrodes.

Experimental Procedures

The diagram of the system configuration for the ABR data collection was illustrated in **Figure 2**. A custom wireless hardware platform was built for high-precision ABR measurements. The hardware platform was made up of a low-noise analog frontend, a high-precision analog-to-digital converter (ADC), and the CC3200 Wi-Fi MCU module. Since the ABR amplitude is rather low, the original analog signal was first amplified by an INA188 instrumentation amplifier (Texas Instruments, Dallas, TX, USA) with a gain of 10 and then processed by a high-pass filter with a cutoff frequency of 100 Hz. The filtered signal was then amplified by an INA141 instrumentation amplifier (Texas Instruments, Dallas, TX, USA) with a fixed gain of 100 and finally amplified by a programmable gain amplifier integrated in ADS1299 (Texas Instruments, Dallas, TX, USA), with a gain of 24. The ADS1299 is an ultra-low-noise, 24-bit simultaneous sampling analog-to-digital converter (ADC) that incorporates all commonly required features for extracranial electroencephalogram (EEG) applications. The amplified analog signal was digitized in the ADS1299 analog frontend at a sampling rate of 16,000 Hz, and the raw data were sent from the CC3200 MCU to the PC through Wi-Fi transmission by Transmission Control Protocol (TCP) packets. The electromagnetic interferences introduced by the Wi-Fi transmission were minimized by the randomization of the TCP packet length, so that the interferences of different Wi-Fi TCP packets would not be synchronized to the stimulus onset, leading to cancelation of each other during the averaging. The performance of the custom ABR data collection platform had been evaluated prior to this study, and the internal noise characteristics were comparable to the commercial SynAmps EEG system (Neuroscan). The wireless raw data were then received by a custom MATLAB (MathWorks, Natick, MA, USA) GUI software platform that was capable of real-time digital filtering, ABR waveform averaging, and noise rejection. The raw data were also stored for further off-line analyses.

During the experiments, the subjects were told to comfortably sit on an adjustable backrest chair inside an electromagnetically shielded room. Then the three skin regions mapped out for

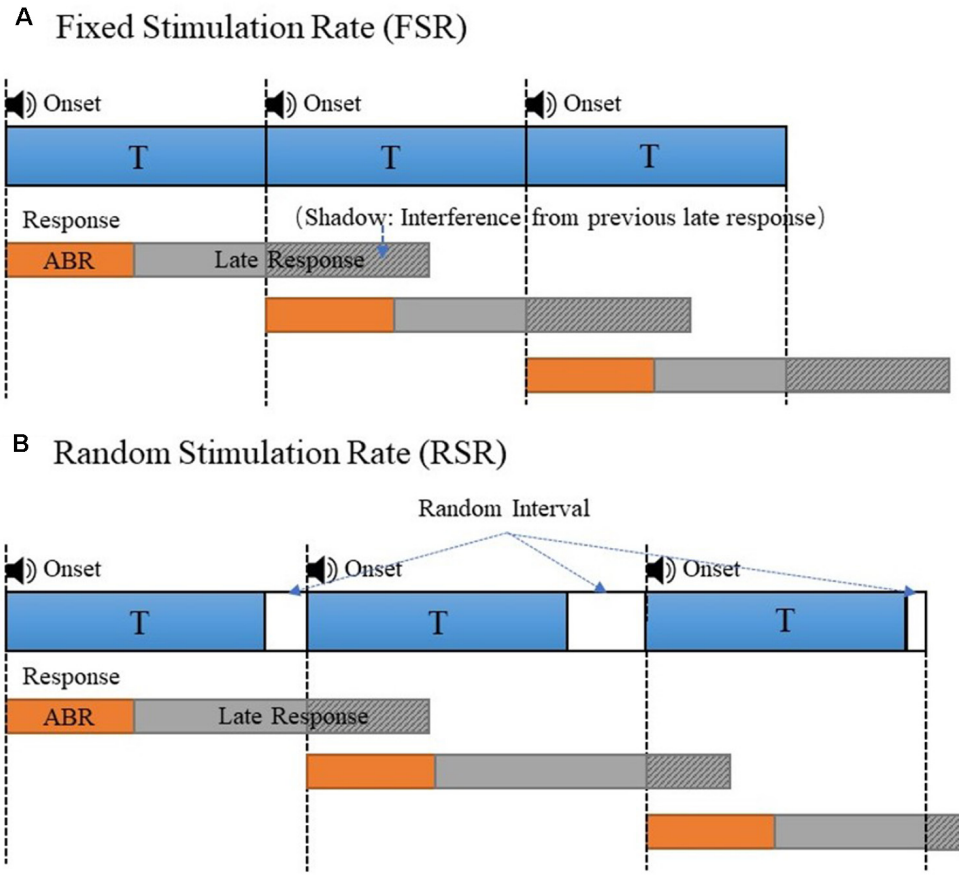

FIGURE 1 | The stimulus presentation comparison of the fixed stimulation rate (FSR; **A**) and random stimulation rate (RSR; **B**) to generate the auditory evoked potential (AEP)s. The overlapped late responses (shadowed area) were synchronous in panel **(A)** and asynchronous in panel **(B)**.

electrode placement were wiped with alcohol pads: the left mastoid, the right mastoid, and the forehead (**Figure 2**). Afterward, the reference (inverting) electrodes were placed over the right mastoid, and the active (non-inverting) electrode was placed over the forehead. Finally, the ground electrode was placed over the left mastoid to minimize the common mode of the reference and active electrodes, using the right leg drive technology incorporated in the ADS1229 chip. The impedance between the skin and the electrodes was screened prior to the experiments to ensure it was below 5 kΩ. The impedance difference between the active and reference electrodes was kept below 1 kΩ for satisfactory common-mode rejection. All the electrodes used in this study were disposable snap electrodes with built-in soft gel.

After all the electrodes were in place, the stimulus was generated from the PC and delivered to the ER-2A earphone that was inserted to the right ear. To eliminate the influence of environmental artifacts on the recorded signals, a foam earplug was inserted into the left ear so that the non-test ear would not have impacts on the ABR results. The earphone wires were kept away from the electrode wires as far as possible to avoid possible electromagnetic interferences when playing the stimuli. For each stimulus condition, the stimulation was repeated 2,000 times for both the FSR and RSR methods. Then the responses of the repeated measures were averaged in reference to the stimulus onset so that the synchronous ABR component could be enhanced while the irrelevant noises would be canceled during the averaging.

In this study, each subject participated in four different experimental sessions to systemically evaluate the performance of the proposed RSR method. A rest time of about 5 min was introduced between two consecutive sessions to prevent the subject from possible fatigue which may degrade the signal quality. In session 1, five trials of the same ABR tests evoked by the RSR method were conducted repeatedly, to examine the test–retest reliability of the proposed method. In this session, a stimulation rate of 20 Hz that was close to the commonly adopted settings in the clinic was used, and the sound intensity of the click stimulus was set to 60 dB SPL. In session 2, the stimulation rate was increased from 20 to 70 Hz with an increment of 10 Hz to explore the performance difference between the FSR and RSR methods under different testing speeds. The stimulus sound intensity was fixed at 60 dB SPL for this session. In session 3, the level of the click stimulus was increased from 50 to 65 dB with a step of 5 dB, and ABR

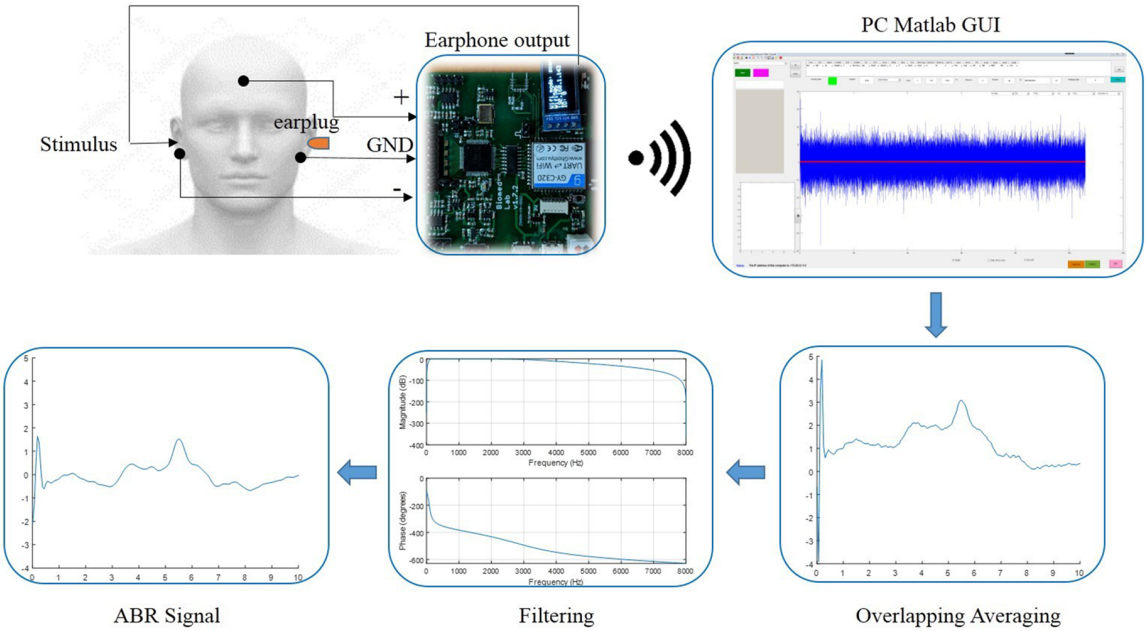

FIGURE 2 | The diagram of the data acquisition system and the configuration of the electrode placement for the auditory brainstem response (ABR) measurements.

signals were measured for both methods. The stimulation rate was constant at 50 Hz. In session 4, the impact of alternate stimulus polarity was investigated for the FSR and RSR methods, with the stimulation rate set to 50 Hz and the stimulus level fixed at 60 dB.

For the data analyses, a digital band-pass filter with cutoff frequencies of 100 and 3,000 Hz was also applied to the raw data to attenuate the out-of-band noises. For the filtered data, an amplitude of 100 μV or greater would be considered as noises (possibly caused by body movements), and the response of the corresponding stimulation would be excluded from the averaging. Then the waveform morphologies and wave I–V latencies of the ABR signals were systemically compared between the FSR and RSR methods, grouped by different stimulus conditions.

EXPERIMENTAL RESULTS

Test–Retest Reliability of the RSR Method

Five repeated trials of the same RSR–ABR measurements were carried out on the same subject, and the temporal waveforms of different trials were compared in **Figure 3** in different colors. The stimulation rate was set to 20 Hz, and the stimulus level was 60 dB SPL. It could be observed from **Figure 3A**, which showed the first 35 ms from the stimulus onset (the beginning of the click stimulus by the earphone), that the peaks and troughs of the five repeated measurements demonstrated good test–retest reliability. It should also be noted that there was a large amplitude of the late responses (such as the peaks of R1 and R2) after the first 10 ms, which could affect the ABR signal if they overlapped with the successive stimulation. The first 10 ms (the actual RSR–ABR signals) of **Figure 3A** was further examined, and the details were shown in **Figure 3B**. It could be seen that all standard peaks from wave I to wave V could be clearly identified for the RSR–ABR waveforms, and the morphologies of all the five trials showed great consistency. Then the latency of each peak of the ABR signals in **Figure 3** was calculated, and the distribution (mean and standard deviation) of the wave latencies from all the five trials was plotted in **Figure 4**. It could be observed that the mean latencies of waves I–V were consistent with related reports in subjects with normal hearing (Nazeri et al., 2016; Cargnelutti et al., 2017; Jiang et al., 2019). The maximum standard deviation of the latencies was as low as 0.2 ms, indicating that all the five trials showed rather consistent waveform morphologies. Similar observations could be found from the results of the other subjects.

Effects of Stimulation Rate

To explore the performance difference between the FSR and RSR methods under different stimulus repetition speeds, the stimulation rate was increased from 20 to 70 Hz, and the comparisons under different speeds (20, 50, and 70 Hz) were shown in **Figure 5** (stimulus level = 60 dB SPL). Generally, the overall ABR waveforms of the two methods were rather consistent across different stimulation rates. However, as compared with the FSR method, the RSR method could achieve better ABR morphologies indicated by clearer waveform differentiation, especially at high stimulation rates (**Figure 5C**). All peaks from waves I to V could be easily recognized for the RSR method regardless of the stimulation rate, whereas waves II and IV were not visible for the FSR method at the stimulation rate of 70 Hz.

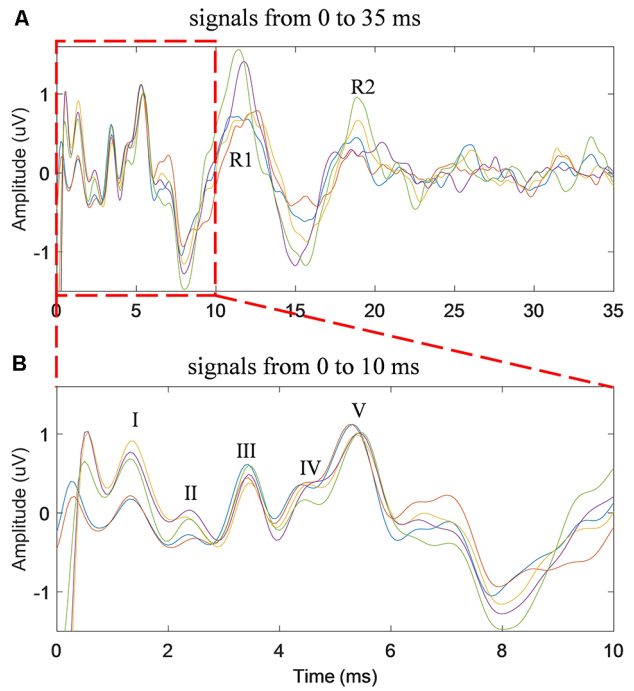

FIGURE 3 | The time waveforms of five repeated trials of the same stimulus condition (stimulation rate of 20 Hz and stimulus level of 60 dB SPL) for the duration of 35 ms **(A)** and 10 ms **(B)** from the stimulus onset.

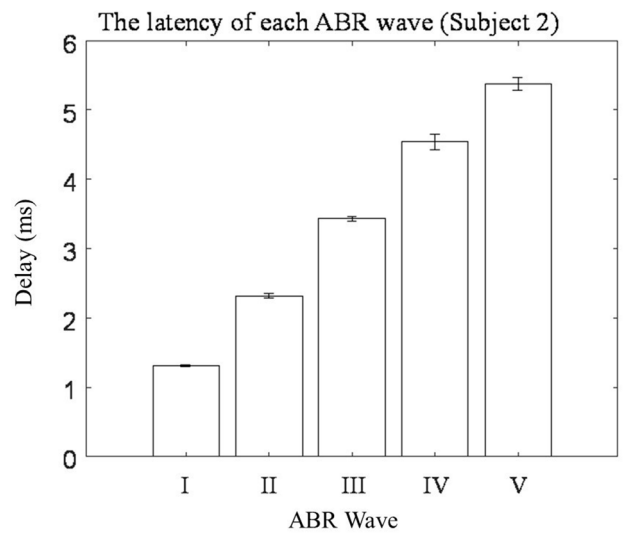

FIGURE 4 | The means and standard deviations of the latencies of wave I–V averaged across all the five repeated trials of ABR tests.

Effects of Stimulus Level

To further examine the performance of the proposed RSR method, the stimulus level was increased from 50 to 65 dB with a step of 5 dB, and the comparisons with the conventional FSR method were shown in **Figure 6** (stimulation rate = 50 Hz). While the waveforms of both methods deteriorated as the stimulus level decreased, the RSR method was less affected by random noises and demonstrated slightly smoother morphologies. The RSR method also showed larger wave V amplitudes at lower stimulus levels of 55 and 50 dB SPL. Similar effects of stimulus level were observed in the results of other subjects.

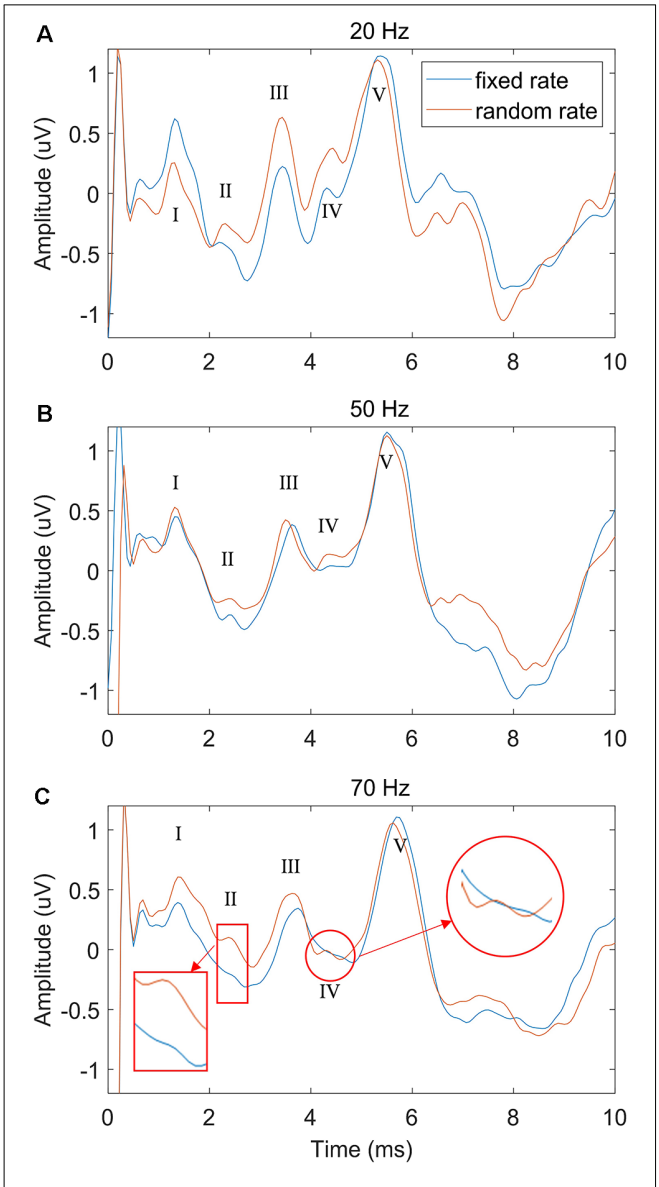

FIGURE 5 | The comparison of ABR waveforms between fixed and random stimulation rates under different repetition speeds of 20 Hz **(A)**, 50 Hz **(B)** and 70 Hz **(C)**, with the stimulation level fixed at 60 dB SPL.

Effects of Stimulus Polarity

In this study, two ways of manipulating the stimulus polarity were performed: non-alternate (only condensation click stimuli were used) and alternate (the condensation and rarefaction clicks were used alternately). The effects of stimulus polarity on the performance of both the FSR and RSR methods were compared in **Figure 7**. It could be observed that large amplitudes of stimulus artifacts were present at the stimulus onset (t = 0 ms, marked by red ovals) for the

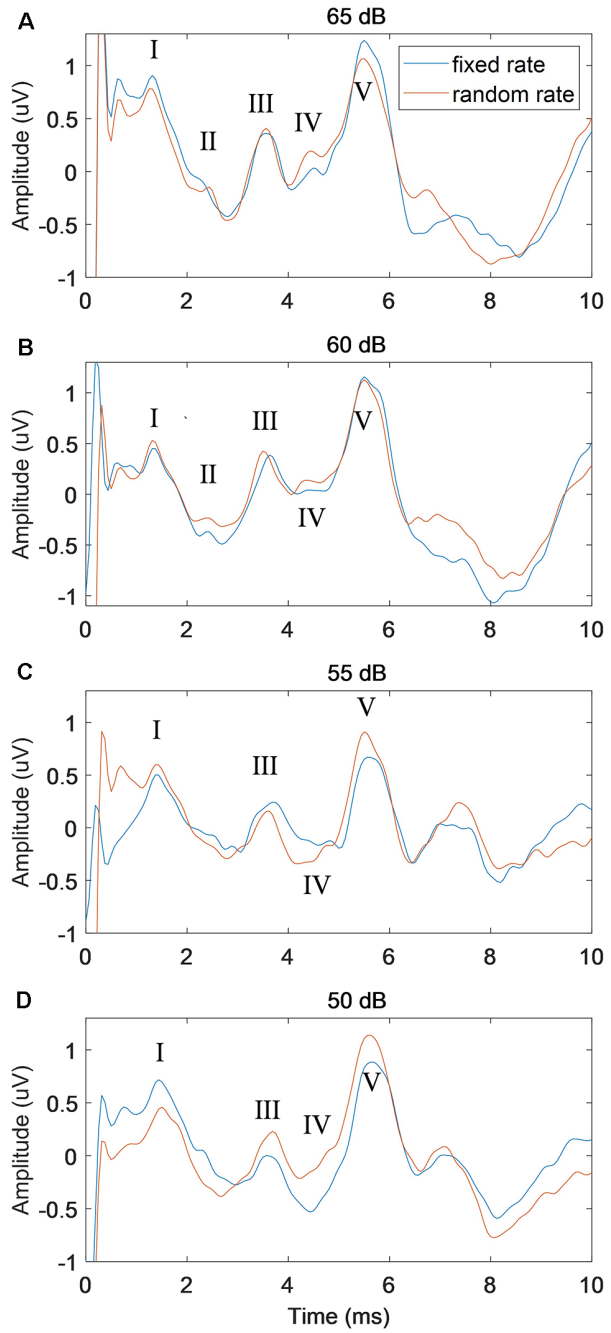

FIGURE 6 | The comparison of ABR waveforms between fixed and random stimulation rates under different stimulus levels of 65 dB **(A)**, 60 dB **(B)**, 55 dB **(C)** and 50 dB **(D)** SPL, with the stimulation rate fixed at 50 Hz.

non-alternate stimulus polarity. In contrast, the alternate stimulus polarity approach could eliminate such stimulus artifacts by canceling the stimulus-related components of the condensation and rarefaction clicks. However, no significant effects of the stimulus polarity on the amplitudes and latencies of waves I to V were observed, for both the FSR and RSR methods.

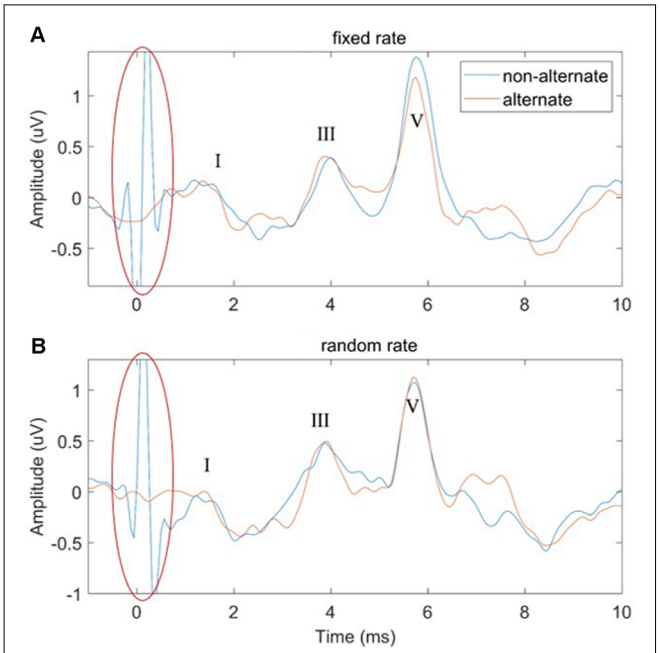

FIGURE 7 | The comparison of ABR waveforms using non-alternate and alternate stimulus polarities for the fixed-rate **(A)** and random-rate **(B)** methods.

Effects of Averaging Times

In order to investigate the speed of convergence (stabilization) of the RSR method, the correlation coefficients between the ABR waveform of increasing average times (from 200 to 1,400 with an increment of 200) and the final ABR waveform (averaged a total of 2,000 times) were calculated for each subject. Then the correlation coefficients of all the subjects were analyzed, with the correlation coefficient statistics (mean and standard deviation) plotted as functions of averaging times (or repeated times) shown in **Figure 8**. The most noteworthy observation was that the convergence speed of the RSR method was significantly faster than that of the conventional FSR method for different stimulus levels, indicating that the proposed RSR method could achieve ABR waveforms similar to the final results (averaged 2,000 times) much earlier. For the same averaging time, the correlation coefficient of the RSR method was also significantly higher. Especially at the stimulus level of 60 dB, the mean correlation coefficient of the RSR method was as high as 0.87 for the averaging time of only 200, whereas the mean correlation coefficient of the FSR method was only 0.72 as a comparison.

DISCUSSION

Although some studies were made on the medical system platform and acquisition system (Han et al., 2019; Pirbhulal et al., 2019; Sun et al., 2020), it helped less with the improvement of ABR morphology. Towards effectively solving the problem on the quality of ABR signals, the RSR method was proposed to improve the morphology and reliability of current ABR

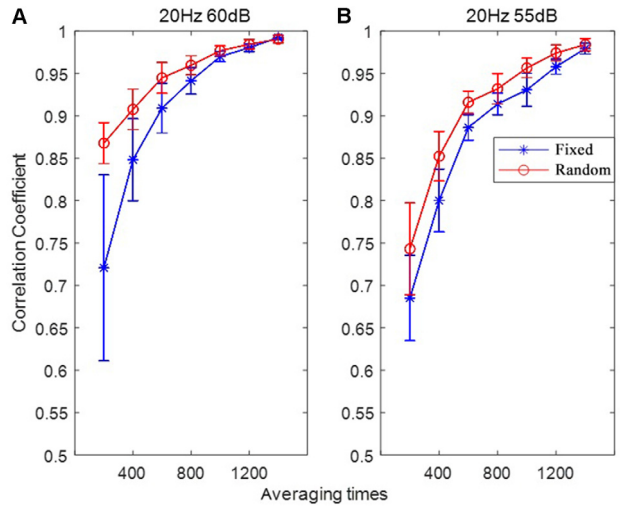

FIGURE 8 | The correlation coefficients between the ABR waveform with varying averaging time and the final ABR waveform (averaged for 2,000 times) under the stimulation conditions of 20 Hz, 60 dB **(A)** and 20 Hz, 55 dB **(B)**.

measurements. By integrating a random interval between two adjacent stimulations, the proposed method could mitigate the interferences arising from the late evoked response of the previous stimulation. The performance of the proposed method was systemically evaluated and compared with the currently used FSR method under various stimulation conditions using a custom wireless high-precision data acquisition platform.

Effects of Stimulation Rate

The stimulation rate has been proven as an important factor that significantly affects the ABR results in the context of clinical hearing loss diagnosis (Schwartz and Morris, 1991; Musiek et al., 1994). In this study, different stimulation rates from 20 to 70 Hz were used to evoke the ABR potentials. The results showed that the ABR waveform differentiations deteriorated as the stimulation rate increased (**Figure 5**), and the findings are consistent with the other studies (Don et al., 1977; Kjær, 1980; Lasky, 1984; Valderrama et al., 2012). Moreover, it was found that at high stimulation rates such as 70 Hz, waves II and IV started to disappear for the conventional FSR method (Alvarez et al., 2010; Valderrama et al., 2012). The reason might be that a large amplitude of late response (10 ms after the stimulus onset, as seen in **Figure 3A**) overlapped with the response of the successive stimulation, given that there was only about 14 ms of time gap between the two stimulus onsets. The previous overlapped late response mixed with the ABR signal of the current stimulation and would be synchronously enhanced after the averaging for the FSR method (**Figure 1A**), leading to significant morphology changes in the ABR measurements. As a comparison, the proposed RSR method could eliminate the synchrony of the overlapped late responses, making them cancel each other during the averaging. Therefore, the waveform differentiation of the RSR method was significantly improved when compared with the FSR method, indicated by the presence of waves II and IV even at a stimulation rate of 70 Hz (**Figure 5C**). The proposed method could make the ABR measurements to be carried out with a faster stimulation rate while maintaining reliable waveform differentiations.

Effects of Stimulus Level

A general observation on the effects of the stimulus level was that better ABR waveform morphologies and shorter wave V latencies could be seen at higher stimulus levels (**Figure 6**), which is consistent with other studies (Serpanos et al., 1997; Louza et al., 2016; Rouillon et al., 2016). **Figure 6** also showed that the proposed RSR method could obtain cleaner ABR signals and better waveform morphologies when compared with the FSR method. The finding might be explained by the cancellation of the overlapped late response of the previous stimulation, as well as other noise sources that were synchronized with the stimulus onset. For low stimulus levels such as 55 and 50 dB, the RSR method could also obtain a larger amplitude of wave V (**Figures 6 C,D**), indicating that the proposed RSR method might be able to measure ABR thresholds at lower stimulus levels and therefore could provide more accurate results for clinical evaluation of hearing functions.

Effects of Stimulus Polarity

In this study, **Figure 7** showed that the large amplitude of stimulus artifacts happening at the stimulus onset could be efficiently eliminated by alternatively changing the polarity of the click stimuli. Similar findings are also reported by other studies when altering the stimulus polarity to measures ABR signals (Gorga et al., 1985; Akhoun et al., 2008; Hornickel et al., 2012; Anderson et al., 2013; Ahadi et al., 2014; Mamo et al., 2016). This can be explained by the linear relationship between the stimulus artifacts and the click polarity. However, no significant effects of the stimulus polarity on the amplitudes and latencies of the waveform peaks were observed for both the FSR and RSR methods. Salt and Thornton (1984) also found that the major component of the ABR was insensitive to stimulus polarity. Regarding the latency of wave V, while some studies reported that there were some differences when changing the stimulus polarity (Borg and Löfqvist, 1981; Hughes et al., 1981; Pijl, 1987), other studies showed that there was no significant difference at all (Rosenhamer et al., 1978; Beattie and Boyd, 1984; Tietze and Pantev, 1986; Kumar et al., 2014). The discrepancy might be attributed to the high sensitivity of ABR signals to various noises. The present study suggests that stimulus polarity is not an important factor when choosing the stimuli in routine clinical ABR measurements.

Effects of Averaging Times

Another important finding of this study is that the convergence speed of the RSR method was significantly faster for different stimulus levels when plotting the correlation coefficients as a function of averaging time (**Figure 8**), indicating that the proposed method could obtain stable ABR waveform earlier than the conventional FSR method. With an averaging time of only 200, the mean correlation coefficient between the current and final ABR waveforms was as high as 0.87, given that over 4,000 averages are usually required to achieve a satisfactory result (Johnson et al., 2008; Hornickel et al., 2009; Skoe and Kraus,

2010; Skoe et al., 2015). The statistics in **Figure 8** suggest that the proposed RSR method is superior in preventing undesired noises from contaminating the target ABR signals so that the waveforms could be stabilized faster than the currently used FSR method. Therefore, the efficiency of current ABR tests could be significantly improved by incorporating the RSR paradigm into clinical settings.

Clinical Implications

In clinical applications, ABR parameters such as the amplitude of peaks, the absolute latency of waves I–V, interpeak latency, and interaural latency are of great importance for the diagnoses of hearing loss and other hearing impairments. However, all these ABR parameters heavily rely on the ABR signal quality and waveform morphologies. Given that the ABR waveform of the conventional fixed-rate method might be affected by interferences originating from the previous late response, the proposed RSR showed great performance in improving the ABR waveform differentiation under different stimulus conditions. Generally, the amplitude and latency of waves I, III, and V and their inter-wave latencies will be taken as the diagnostic parameters for hearing impairments. However, the proposed RSR method could help to identify clear waves II and IV, whose amplitudes and latencies also provide rather useful information for clinical diagnoses. For example, the inter-wave latency of waves III and IV reflects axonal conduction time, and the interval of wave IV and V reflects a synaptic delay (De Vries and Glass, 2019). Lee et al. (2018) reported that the amplitude of wave II could be considered a supplementary indicator to help with the diagnosis of vestibular paroxysmia.

Compared with Valderrama's iterative-RSA method that might involve complex deconvolution algorithms and empirically chosen controlling factors (Valderrama et al., 2014), the implementation strategy of the proposed method is much simpler, with only minimal changes to the stimulus presentation of the current commercial systems. The easy implementation of the proposed method would help to greatly reduce the cost while achieving significant improvement in ABR signal qualities, making it rather useful in medical application scenarios such as intraoperative monitoring during surgery, auditory threshold estimation, and newborn hearing screening.

CONCLUSION

In this study, an RSR method was proposed, and the performance on improving the morphology and reliability of ABR signals was systemically investigated under different stimulus conditions. The results showed that the RSR method demonstrated great test–retest reliability in repeated measurements. By canceling the interferences of the late response from the previous stimulation, it could also achieve better ABR morphologies indicated by clearer waveform differentiation under different stimulation rates and stimulus levels. The RSR method could obtain satisfactory results significantly faster than the conventional FSR method, and it could help to greatly improve the efficiency of current ABR measurements. The proposed RSR method may provide a candidate tool that would aid accurate and efficient diagnoses of hearing impairment in clinical settings. The approach of obtaining reliably evoked potentials from the brain might also be helpful for applications such as brain–computer interface and intelligent control of robotic systems.

DATA AVAILABILITY STATEMENT

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

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-190615-H0352). The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

XinW, MZ, and SC contributed to the conception and design of the study. OS, XiaW, and HZ conducted experiments. YL and WW analyzed the data. SM contributed to the choice of the electrodes. XinW wrote the first draft of the manuscript. JY, MW, and GL contributed to manuscript revision. All authors read and approved the submitted manuscript.

FUNDING

This work was supported in part by a Shenzhen Governmental Basic Research Grant (#JCYJ20180507182241622), the National Natural Science Foundation of China (#61771462 and #61901464), Science and Technology Program of Guangzhou (#201803010093), and Science and Technology Planning Project of Guangdong Province (#2019A050510033).

REFERENCES

Ahadi, M., Pourbakht, A., Jafari, A. H., and Jalaie, S. (2014). Effects of stimulus presentation mode and subcortical laterality in speech-evoked auditory brainstem responses. *Int. J. Audiol.* 53, 243–249. doi: 10.3109/14992027.2013.866281

Aimoni, C., Ciorba, A., Bovo, R., Trevisi, P., Busi, M., and Martini, A. (2010). Hearing threshold assessment in young children with electrocochleography (EcochG) and auditory brainstem responses (ABR): experience at the University Hospital of Ferrara. *Auris Nasus Larynx* 37, 553–557. doi: 10.1016/j.anl.2010.02.002

Akhoun, I., Gallégo, S., Moulin, A., Ménard, M., Veuillet, E., Berger-Vachon, C., et al. (2008). The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme/ba/in normal-hearing adults. *Clin. Neurophysiol.* 119, 922–933. doi: 10.1016/j.clinph.2007.12.010

Alvarez, I., Valderrama, J. T., DeLaTorre, A., Segura, J. C., Sainz, M., and Vargas, J. L. (2010). "Reducción del tiempo de exploración de potenciales evocados auditivos del tronco cerebral mediante estimulación aleatorizada (Brainstem auditory evoked potentials time reduction through randomized stimulation)", in *Oral Presentation at the XXV Unión Cientifica Internacional De Radio (URSI) National Symposium*, Bilbao, Spain.

Alwan, A. (2012). *Implementation of Wavelet-Kalman Filtering Technique for Auditory Brainstem Response*. Linkoping, Sweden: Linkoping University, Department of Electrical Engineering.

Anderson, S., Parbery-Clark, A., White-Schwoch, T., and Kraus, N. (2013). Auditory brainstem response to complex sounds predicts self-reported speech-in-noise performance. *J. Speech Lang. Hear. Res.* 56, 31–43. doi: 10.1044/1092-4388(2012/12-0043)

Avan, P., and Bonfils, P. (1997). *Exploration Fonctionnelle Objective Des Voies Auditives*. Cachan, France: Éditions Médicales Internationales, 113–134.

Beattie, R. C., and Boyd, R. (1984). Effects of click duration on the latency of the early evoked response. *J. Speech Hear. Res.* 27, 70–76. doi: 10.1044/jshr.2701.70

Borg, E., and Löfqvist, L. (1981). Brainstem response (ABR) to rarefaction and condensation clicks in normal hearing and steep high-frequency hearing loss. *Scand. Audiol.* 13, 99–101.

Canale, A., Dagna, F., Lacilla, M., Piumetto, E., and Albera, R. (2020). Relationship between pure tone audiometry and tone burst auditory brainstem response at low frequencies gated with blackman window. *Eur. Arch. Otorhinolaryngol.* 269, 781–785. doi: 10.1007/s00405-011-1723-7

Cargnelutti, M., Cóser, P. L., and Biaggio, E. P. V. (2017). LS CE-Chirp® vs. click in the neuroaudiological diagnosis by ABR. *Braz. J. Otorhinolaryngol.* 83, 313–317. doi: 10.1016/j.bjorl.2016.04.018

Ceylan, S., Gümüşgün, A., and Feratlar, F. (2018). Comparison of CE-Chirp ABR and click ABR methods in patients with bilateral sensorineural hearing loss. *ENT Updates* 8, 27–32. doi: 10.2399/jmu.2018001009

de Oliveira, J. M., Munoz, R., Ribeiro, S., Wu, W., and De Albuquerque, V. H. C. (2019). REHAB FUN: an assistive technology in neurological motor disorders rehabilitation of children with cerebral palsy. *Neural Comput. Appl.* 1–14. doi: 10.1007/s00521-019-04059-2 [Epub ahead of Print].

De Vries, L. S., and Glass, H. C. (2019). *Neonatal Neurology: Handbook of Clinical Neurology Series.* Amsterdam: Elsevier.

Don, M., Allen, A. R., and Starr, A. (1977). Effect of click rate on the latency of auditory brain stem responses in humans. *Annals of Otology, Rhinology & Laryngology* 86, 186–195. doi: 10.1177/000348947708600209

Eggermont, J. J., and Moore, J. K. (2012). "Morphological and functional development of the auditory nervous system," in *Human Auditory Development*, eds L. Werner, R. Fay and A. Popper (New York, NY: Springer), 61–105.

Elzouki, A., Harfi, H. A., Nazer, H., Stapleton, F. B., Oh, W., and Whitley, R. J. (2012). *Textbook of Clinical Pediatrics.* New York, NY: Springer Science & Business Media.

Galambos, R., and Hecox, K. (1978). Clinical applications of the auditory brain stem response. *Otolaryngol. Clin. North Am.* 11, 709–722.

Gorga, M., Abbas, P., and Worthington, D. (1985). "Stimulus calibration in ABR measurements," in *Auditory Brainstem Response*, ed. J. Jacobsen (San Diego, CA: College Hill Press), 49–62.

Han, T., Zhang, L., Pirbhulal, S., Wu, W., and De Albuquerque, V. H. C. (2019). A novel cluster head selection technique for edge-computing based IoMT systems. *Compu. Net.* 158, 114–122. doi: 10.1016/j.comnet.2019.04.021

Hoda, A.-M., Sayed, E., and Mahran, S. (2019). Auditory brainstem response to chirp stimulus in children with moderate and severe sensorineural hearing loss. *Egypt. J. Otolaryngol.* 35:322. doi: 10.4103/ejo.ejo_25_18

Hornickel, J., Knowles, E., and Kraus, N. (2012). Test-retest consistency of speech-evoked auditory brainstem responses in typically-developing children. *Hear. Res.* 284, 52–58. doi: 10.1016/j.heares.2011.12.005

Hornickel, J., Skoe, E., and Kraus, N. (2009). Subcortical laterality of speech encoding. *Audiol. Neurootol.* 14, 198–207 doi: 10.1159/000188533

Hughes, J. R., Fino, J., and Gagnon, L. (1981). The importance of phase of stimulus and the reference recording electrode in brain stem auditory evoked potentials. *Electroencephalogr. Clin. Neurophysiol.* 51, 611–623. doi: 10.1016/0013-4694(81)90205-4

James, S. L., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., et al. (2018). Global, regional and national incidence, prevalence and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017. *Lancet* 392, 1789–1858. doi: 10.1016/S0140-6736(18)32279-7

Jiang, Y., Samuel, O., Liu, X., Wang, X., Idowu, O., Li, P., et al. (2018). Effective biopotential signal acquisition: comparison of different shielded drive technologies. *Appl. Sci.* 8, 276–295. doi: 10.3390/app8020276

Jiang, Y., Wang, D., Liu, Z., Tan, J., Samuel, O. W., Deng, H., et al. (2019). "Comparing auditory brainstem responses evoked by click and sweep-tone in normal-hearing adults," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (Berlin: IEEE), 5237–5240.

Johnson, K. L., Nicol, T., Zecker, S. G., Bradlow, A. R., Skoe, E., and Kraus, N. (2008). Brainstem encoding of voiced consonant-vowel stop syllables. *Clin. Neurophysiol.* 119, 2623–2635 doi: 10.1016/j.clinph.2008.07.277

Kjær, M. (1980). Brain stem auditory and visual evoked potentials in multiple sclerosis. *Acta Neurol. Scand.* 62, 14–19. doi: 10.1111/j.1600-0404.1980.tb02999.x

Kumar, K., Bhat, J. S., D'Costa, P. E., Srivastava, M., and Kalaiah, M. K. (2014). Effect of stimulus polarity on speech evoked auditory brainstem response. *Audiol. Res.* 3:e8. doi: 10.4081/audiores.2013.e8

Lasky, R. E. (1984). A developmental study on the effect of stimulus rate on the auditory evoked brain-stem response. *Clin. Neurophysiol.* 59, 411–419. doi: 10.1016/0168-5597(84)90042-x

Lee, J. H., Hong, S.-K., Kim, H.-J., and Lee, H.-J. (2018). Is the auditory brainstem response diagnostic for vestibular paroxysmia? *Res. Vestibul. Sci.* 17, 55–59.doi: 10.21790/rvs.2018.17.2.55

Louza, J., Polterauer, D., Wittlinger, N., Muzaini, H. A., Scheckinger, S., Hempel, M., et al. (2016). Threshold changes of ABR results in toddlers and children. *Int. J. Pediatr. Otorhinolaryngol.* 85, 120–127. doi: 10.1016/j.ijporl.2016.03.009

Lu, T.-M., Wu, F.-W., Chang, H., and Lin, H.-C. (2017). Using click-evoked auditory brainstem response thresholds in infants to estimate the corresponding pure-tone audiometry thresholds in children referred from UNHS. *Int. J. Pediatr. Otorhinolaryngol.* 95, 57–62. doi: 10.1016/j.ijporl.2017.02.004

Mamo, S. K., Grose, J. H., and Buss, E. (2016). Speech-evoked ABR: effects of age and simulated neural temporal jitter. *Hear. Res.* 333, 201–209. doi: 10.1016/j.heares.2015.09.005

Musiek, F., Borenstien, S., Hall, J., and Schwaber, M. (1994). *Auditory Brainstem Response: Neurodiagnostic and Intraoperative Applications. Handbook of Clinical Audiology* (Baltimore, MD: Williams and Wilkins), 351–374.

Nazeri, A. R., Moosavi, A., Lotfi, Y., and Bakhshi, E. (2016). Comparison of auditory evoked potentials between younger and older-adults. *J. Otorhinolaryngol. Facial Plast. Surg.* 2, 29–36. doi: 10.22037/orlfps.v2i2.12044

Olusanya, B. O., Neumann, K. J., and Saunders, J. E. (2014). The global burden of disabling hearing impairment: a call to action. *Bull. World Health Organ.* 92, 367–373. doi: 10.2471/blt.13.128728

Pijl, S. (1987). Effects of click polarity on ABR peak latency and morphology in a clinical population. *J. Otolaryngol.* 16, 89–96.

Pirbhulal, S., Samuel, O. W., Wu, W., Sangaiah, A. K., and Li, G. (2019). A joint resource-aware and medical data security framework for wearable healthcare systems. *Future Gene. Comput. Syst.* 95, 382–391. doi: 10.1016/j.future.2019.01.008

Pirbhulal, S., Zhang, H., Wu, W., Mukhopadhyay, S. C., and Zhang, Y.-T. (2018). Heartbeats based biometric random binary sequences generation to secure wireless body sensor networks. *IEEE Trans. Biomed. Eng.* 65, 2751–2759. doi: 10.1109/tbme.2018.2815155

Rosenhamer, H., Lindström, B., and Lundborg, T. (1978). On the use of click-evoked electric brainstem responses in audiological diagnosis: I. *Scand. Audiol.* 7, 193–205. doi: 10.3109/01050397809076287

Rouillon, I., Parodi, M., Denoyelle, F., and Loundon, N. (2016). How to perform ABR in young children. *Eur. Ann. Otorhinolaryngol. Head Neck Dis.* 133, 431–435. doi: 10.1016/j.anorl.2016.05.004

Salt, A., and Thornton, A. (1984). The effects of stimulus rise-time and polarity on the auditory brainstem responses. *Scand. Audiol.* 13, 119–127. doi: 10.3109/01050398409043050

Samuel, O. W., Geng, Y., Li, X., and Li, G. (2017a). Towards efficient decoding of multiple classes of motor imagery limb movements based on EEG spectral and time domain descriptors. *J. Med. Syst.* 41:194. doi: 10.1007/s10916-017-0843-z

Samuel, O. W., Li, X., Geng, Y., Feng, P., Chen, S., and Li, G. (2017b). "Motor imagery classification of upper limb movements based on spectral domain features of EEG patterns," in *2017 39th Annual International Conference of the*

IEEE Engineering in Medicine and Biology Society (EMBC), Jeju Island, South Korea, 2976–2979.

Schwartz, D., and Morris, M. (1991). Strategies for optimizing the detection of neuropathology from the auditory brainstem response. *Diagnost. Audiol.*,141–160.

Serpanos, Y. C., O–malley, H., and Gravel, J. S. (1997). The relationship between loudness intensity functions and the click-ABR wave V latency. *Ear Hear.* 18, 409–419. doi: 10.1097/00003446-199710000-00006

Skoe, E., and Kraus, N. (2010). Auditory brainstem response to complex sounds: a tutorial. *Ear Hear.* 31:302. doi: 10.1097/aud.0b013e3181cdb272

Skoe, E., Krizman, J., Anderson, S., and Kraus, N. (2015). Stability and plasticity of auditory brainstem function across the lifespan. *Cereb. Cortex* 25, 1415–1426. doi: 10.1093/cercor/bht311

Sun, X., Wang, S., Xia, Y., and Zheng, W. (2020). ''Predictive-trend-aware composition of web services with time-varying quality-of-service,'' *IEEE Access*, Voume 8 (Piscataway, NJ: IEEE), 1910–1921.

Talaat, H. S., Hammad, A., and El Abedein, A. M. Z. (2020). Hearing threshold evaluation in children using narrow band chirp auditory brainstem response and tone burst auditory brainstem response. *Int. J. Otolaryngol. Head & Neck Surg.* 9:30 doi: 10.4236/ijohns.2020.91005

Tietze, G., and Pantev, C. (1986). Comparison between auditory brain stem responses evoked by rarefaction and condensation step functions and clicks. *Audiology* 25, 44–53. doi: 10.3109/00206098609078368

Valderrama, J. T., Alvarez, I., de la Torre, A., Carlos Segura, J., Sainz, M., and Luis Vargas, J. (2012). Recording of auditory brainstem response at high stimulation rates using randomized stimulation and averaging. *J. Acoust. Soc. Am.* 132, 3856–3865. doi: 10.1121/1.4764511

Valderrama, J. T., De La Torre, A., Alvarez, I., Segura, J. C., Thornton, A. R. D., Sainz, M., et al. (2014). A study of adaptation mechanisms based on ABR recorded at high stimulation rate. *Clin. Neurophysiol.* 125, 805–813. doi: 10.1016/j.clinph.2013.06.190

Vos, T., Barber, R. M., Bell, B., Bertozzi-Villa, A., Biryukov, S., Bolliger, I., et al. (2015). Global, regional and national incidence, prevalence and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the global burden of disease study 2013. *Lancet* 386, 743–800. doi: 10.1016/S0140-6736(15)60692-4

Wong, P. K., and Bickford, R. G. (1980). Brain stem auditory evoked potentials: the use of noise estimate. *Clin. Neurophysiol.* 50, 25–34. doi: 10.1016/0013-4694(80)90320-x

Wu, W., Pirbhulal, S., and Li, G. (2018). Adaptive computing-based biometric security for intelligent medical applications. *Neural Comput. Appl.*, 1–10. doi: 10.1007/s00521-018-3855-9

World Health Organization (2011). *The Global Burden of Disease: 2004 Update.2008*. Geneva: World Health Organization

World Health Organization (2015). *Deafness and Hearing Loss, Fact Sheet Ns 300*. Geneva: WHO Media Centre.

Xie, L., Wang, M., Liao, T., Tan, S., Sun, K., Li, H., et al. (2018). The characterization of auditory brainstem response (ABR) waveforms: a study in tree shrews (*Tupaia belangeri*). *J. Otol.* 13, 85–91. doi: 10.1016/j.joto.2018.05.004

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

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 07 April 2020
doi: 10.3389/fnhum.2020.00070

Assessing the Effectiveness of Automated Emotion Recognition in Adults and Children for Clinical Investigation

Maria Flynn[1], Dimitris Effraimidis[2], Anastassia Angelopoulou[2], Epaminondas Kapetanios[2], David Williams[1], Jude Hemanth[3] and Tony Towell[1]*

[1] School of Social Sciences, University of Westminster, London, United Kingdom, [2] School of Computer Science and Engineering, University of Westminster, London, United Kingdom, [3] ECE Department, Karunya Institute of Technology and Sciences, Coimbatore, India

OPEN ACCESS

Edited by:
Victor Hugo C. de Albuquerque, University of Fortaleza, Brazil

Reviewed by:
Oana Geman, Ştefan cel Mare University of Suceava, Romania
Fuqian Shi, University of Central Florida, United States
Bogdan Patrut, Alexandru Ioan Cuza University, Romania
Utku Kose, Süleyman Demirel University, Turkey

***Correspondence:**
Jude Hemanth
judehemanth@karunya.edu

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

Received: *16 January 2020*
Accepted: *17 February 2020*
Published: *07 April 2020*

Citation:
Flynn M, Effraimidis D, Angelopoulou A, Kapetanios E, Williams D, Hemanth J and Towell T (2020) Assessing the Effectiveness of Automated Emotion Recognition in Adults and Children for Clinical Investigation. Front. Hum. Neurosci. 14:70. doi: 10.3389/fnhum.2020.00070

Recent success stories in automated object or face recognition, partly fuelled by deep learning artificial neural network (ANN) architectures, have led to the advancement of biometric research platforms and, to some extent, the resurrection of Artificial Intelligence (AI). In line with this general trend, inter-disciplinary approaches have been taken to automate the recognition of emotions in adults or children for the benefit of various applications, such as identification of children's emotions prior to a clinical investigation. Within this context, it turns out that automating emotion recognition is far from being straightforward, with several challenges arising for both science (e.g., methodology underpinned by psychology) and technology (e.g., the iMotions biometric research platform). In this paper, we present a methodology and experiment and some interesting findings, which raise the following research questions for the recognition of emotions and attention in humans: (a) the adequacy of well-established techniques such as the International Affective Picture System (IAPS), (b) the adequacy of state-of-the-art biometric research platforms, (c) the extent to which emotional responses may be different in children and adults. Our findings and first attempts to answer some of these research questions are based on a mixed sample of adults and children who took part in the experiment, resulting in a statistical analysis of numerous variables. These are related to both automatically and interactively captured responses of participants to a sample of IAPS pictures.

Keywords: emotion, brain, artificial neural network, computing, clinical investigation

1. INTRODUCTION

Emotions are the essence of what makes us human. Emotional response can be measured by at least three different systems: affective reports, physiological reactivity, and overt behavioral acts (Lang, 1969). One of the strongest indicators for our emotions has always been considered our face. Cross-cultural studies suggest that there is a set of universal basic emotions that can be recognized from facial expressions, including anger, disgust, fear, sadness, and enjoyment (Ekman, 1993). Facial expressions are a strong correlate of emotion, and it has been shown that almost everyone can produce and recognize facial expressions (Ekman and Friesen, 1978; Ekman, 2016). Consequently, previous studies have investigated emotional reactions using affective pictures to elicit emotional experience in adults (Greenwald et al., 1989) and in children (McManis et al., 2001).

Prominent position in these studies was taken by the International Affective Picture System (IAPS) (Lang et al., 1997), which provides a set of normative emotional stimuli for experimental investigations of emotion and attention. When used in combination with tools for the collection of subjective affective ratings such as the Self-Assessment Manikin (Bradley and Lang, 1994) or the Affective Slider (Betella and Verschure, 2016), which are non-verbal assessment techniques that directly measure the pleasure and arousal associated with a wide variety of stimuli, emotional affect can be measured. Furthermore, skin conductance is also a sensitive autonomic measure of emotional arousal (Boucsein et al., 2012). The higher the arousal, the higher the skin conductance for both positive ("happy" or "joyful") and negative ("threatening" or "saddening") stimuli. Consequently, biometric research platforms have emerged specializing in computer vision and machine learning techniques, which enable reliable, valid measurement of emotion-related facial expressions from real-time non-invasive sensors (Sikka et al., 2015). Combining computer machine learning techniques that measure facial expressions with skin conductance responses and self-report may provide useful insight into emotional states.

Despite all these technological advancements, there is currently an ongoing lively debate about the effectiveness of automated emotion recognition approaches. For instance, there seems to be a paradigm shift from the basic emotion perspective to an appraisal perspective to find the appropriate theory integration in the area of automated facial emotion classification. In general, the criticism of the basic emotion perspective argues that, although automated facial expression analysis may classify basic emotional expression categories, it might not ultimately measure emotional states.

The fact that automated facial expression analysis relies on the assumption that there is coherence between emotion and facial expressions (Bonanno and Keltner, 2004; Reisenzein et al., 2013) limits the interpretation of data generated by automated facial expression analysis and throws into question the generalization of automated emotion classification (Wolf, 2015). Furthermore, some researchers argue that inferences based on data generated by automated facial expression analysis should build upon emotion theories that go beyond the basic emotion perspective, adopt an appraisal perspective, and allow more flexibility to consider different contexts.

Further to this criticism, relying on machine learning techniques and algorithms also raises the question of whether the algorithmic design and implementation introduced is transparent and also discrimination- and fairness-aware. It is only then that classifications or predictions, such as those imposed by the foreseen recognition of emotions in children prior to clinical investigation, are trustworthy and not subject to bias. Generally speaking, there are two sources of bias to be prevented: (a) *data sources and input*, (b) *algorithms* (Hajian and Bonchi, 2016).

Given this context, this paper contributes to the lively debate and criticism surrounding the effectiveness of automated emotion recognition approaches. In particular, it presents the results and interesting findings from a study and experiment that set out to determine how children and adults may respond to emotional stimuli and whether such emotions can be adequately captured and analyzed by state-of-the-art biometric research platforms. The study has the potential to advance our ability to identify children in a hospital environment who are very anxious, scared, or upset.

This paper, however, focuses on the observed discrepancies, under specific circumstances, in expected (e.g., IAPS) and observed (e.g., biometric research platform, subjective classification) emotional responses, which may further help in identifying the root for the emergence of such a criticism against automated emotion recognition approaches, in general, and those based on facial recognition, in particular. Hence, the rest of the paper is structured as follows. The state of the art is reviewed in section 2. Section 3 presents our methodology and experimental setup, where materials are thoroughly explained by giving an overview and then focusing on each piece of the system. Then, the first statistical results are presented in sections 4.1 (adults) and 4.2 (children). Finally, we conclude by also considering the outlook for the future.

2. RELATED WORK

This section provides a brief overview of the literature on the attempts to classify emotions and of the development of affective computer systems in relation to facial expression recognition and other computer-based systems developed to recognize human emotions.

Emotion refers to a shaking of the organism as a response to a particular stimulus (person, situation, or event), which is generalized and occupies the person as a whole. Usually, it is very brief in duration, which makes it different than Mood. Mood is a feeling that tends to be less intense than emotion and often lacks a contextual stimulus (Weiss and Cropanzano, 1996; Feidakis et al., 2011). Both emotions and moods are encompassed under the umbrella of "Affect," which is a generic term that covers a broad range of feelings that people experience (George, 1996). Affective computing is the set of techniques aimed at performing affect recognition from data in different modalities and by using multiple sensors in order to increase the reliability of estimates. Affective computing involves two areas: emotion synthesis, which is used to artificially imitate some of the physical or behavioral expressions associated with affective states, and emotion analysis, which is often employed in decision making for interactive systems (Shen et al., 2009; Poria et al., 2017). In this paper, we discuss and implement both kinds of affects by specifically using well-established datasets such as the IAPS for emotion evocation and stimulus and facial recognition software for identifying and analyzing emotions in an adult and child population.

Before discussing the methodology of our affective system, we introduce the literature on the different categorizations of emotions and the state-of-the-art of affective systems in relation to protective groups. In the last several decades, psychologists have categorized emotions with two fundamental viewpoints: (1) discrete categories, and (2) emotions grouped on a dimensional basis. In the first category, all humans

are believed to have an essential set of basic emotions that are distinguishable by an individual's facial expression and biological processes (Colombetti, 1995). Most emotion experts think there are a few "basic emotions," although they do not all agree on what they are or why they are basic (Ortony et al., 1987). Ekman (2003) and Ekman and Friesen (1978) has been very influential with his studies of facial expressions. He developed a list of basic emotions that are universal, not culturally determined. These basic emotions are "anger, disgust, fear, happiness, sadness, and surprise" (Ekman and Friesen, 1978). A few tentative efforts to detect non-basic affective states, such as "fatigue, anxiety, satisfaction, confusion, or frustration" have also been made (Dalgleish and Power, 1999; Prinz, 2004). In the second category, researchers define emotions according to one or more dimensions. Most of these models integrate valence and arousal or intensity dimensions, as they propose that a common and interconnected neurophysiological system is responsible for all affective states (Rubin and Talerico, 2009). Russell 1980; 2003 circumplex model of affect was developed on the basis that affective states arise from cognitive interpretations of core neural sensations that are the product of two independent neuro-physiological systems. The model suggests that emotions are distributed in a two-dimensional circular space, valence and activation (or arousal). Valence represents the horizontal axis. It can be pleasant (positive) such as happiness or joy, or it can be unpleasant (negative), such as anxiety, anger, or boredom. Researchers have criticized two-dimensional models as being too simplistic. Recent evidence suggests there should be a fourth dimension (Fontaine et al., 2007). Watson and Tellegen (1985) changed the orientation and proposed four dimensions: "pleasantness, engagement, positive, and negative affect." Fontaine et al. (2007) reported consistent results from various cultures where a set of four dimensions is found in user studies, namely "valence, potency, arousal, [and] unpredictability." Plutchik (1980, 2003) proposed a cone-shaped model with intensity of emotional experience represented by depth, similarity by nearness, and four pairs of opposites, all represented by color-coded segments.

As we have seen from the above categorization of emotions, Ekman (2003); Ekman and Friesen (1978), a pioneer in the visual modality analysis of emotions, referred to facial expressions as primary cues for understanding emotions and sentiments. Facial expressions are a gateway into the human mind, emotion, and identity and, along with textual data, can provide important cues to better identify true affective states in the participants (Taggart et al., 2016; Kim et al., 2018). It can be crucial to understand facial characteristics when working with patients, especially patients who are unable to communicate in other ways, for example, when trying to assess emotions in children unable to self-report information. This is particularly true when the children have multi-systemic problems and may be dysmorphic, making interpretation of facial expressions even more difficult.

Therefore, in clinical environments, assessments of the child's emotional state are typically made by clinical staff or family members. However, in some instances, staff may have difficulty in accurately estimating children's emotional states, and family members/carers may not always be available. In such cases, automated systems based on computer vision and machine learning techniques that can reliably process and analyze valid measurements of emotion-related facial expressions without using invasive sensors can play a crucial role in diagnostic cases such as autism. Due to the nature of these studies, which have very detailed ethical requirements and require access to data on protected groups, only a handful of studies have examined the efficacy of automated systems in detecting emotional expressions in individuals from protected groups in order to assist and define protocols for better therapeutic treatments. Trevisan et al. (2016) used facial expression analysis technology to determine how children with and without autism spectrum disorder (ASD) may differentially produce facial expressions in response to emotional stimuli and whether alexithymia may contribute to diminished facial expressions. Xefteris et al. (2016) developed a methodology for emotion recognition using facial expressions as indicators to evaluate the overall health status of subjects suffering from neurodegenerative diseases (e.g., Mild Cognitive Impairments, Alzheimer's, dementia). Leo et al. (2015) used machine learning strategies based on facial expressions during robot-child user interaction to evaluate the behaviors of children who belong to the ASD group for the development of better therapeutic protocols. Suzan and Mariofanna (2016) used computer vision and machine learning methods such as active shape models (ASM) and Support Vector Machine (SVM) to recognize facial expressions in children with ASD during playtime. Kunz et al. (2017) used an interdisciplinary approach of human observers and video-based pain detection systems that analyzes facial expressions to identify pain in people with dementia and ensure effective treatment and ongoing care.

In addition, studies using physiological signals to recognize emotional states such as electroencephalogram (EEG)-based brain-computer interface systems (BCI) are also providing interesting results, and there is promise for use in a number of real-world applications. Huang et al. (2019) showed participants video clips with negative and positive valence while recording EEG. The EEG-based BCI system successfully induced and recognized positive and negative emotional states in patients with Disorders of Consciousness. Hou and Chen (2019) presented a system for characterizing emotions using EEG signals, where four classes of emotions in particular (i.e., happy, sad, calm, and angry) could be distinguished. They induced these emotions by musical stimuli (using 20 music passages in each music emotion group) and recorded the EEG signals of the subjects using 12 electrodes. Guan et al. (2019) proposed a novel classification framework using a decision tree (DT) classifier to distinguish between multiclass motor imagery (MI) for BCI. Their proposed data reduction method performed better when compared to state-of-the-art semisupervised joint mutual information (semi-JMI) and general discriminant analysis (GDA) methods. Fernàndez-Rodríguez et al. (2019) used different sets of flashing stimuli in a number of participants in order to assess the effect of the emotional stimuli in these images by using a P300 brain-computer interface (BCI) speller. Finally, it has been demonstrated in various studies (Acharya et al., 2018; Jahmunah et al., 2019) that EEG signals are commonly used to detect brain diseases such as depression and schizophrenia.

All of the above models of emotions are very important when designing and informing the development of affective systems. Dimensional models have been used by various researchers, mainly because they provide a way of describing a wide range of emotional states that occur in real life. "Basic" emotional models were very influential in early human-computer interaction studies. When all these emotional models are put into a computational framework where programmers and developers map aspects of emotion to aspects of the system, different models have different pros and cons (Bosse et al., 2010). In our study, we have used basic emotional models in the categorical data and dimensional models for automatically measuring and analyzing the emotions of related behaviors.

3. METHODOLOGY AND EXPERIMENTAL DESIGN

In the following sections, two experiments will be described, one with adult participants and one with child participants. Studies such as Mikels et al. (2005), which aimed to provide categorical data for the IAPS dataset, have shown that some pictures were rated as evoking a combination of emotions, for example, anger, fear, and disgust. Studies such as Barke et al. (2012) have shown that there are sex differences and cross-cultural differences in the rating and categorization of a subset of IAPS pictures, with women tending to rate negative pictures more negatively and with higher arousal ratings than men, and they established valence and arousal norms for a German sample, suggesting that country and sex-specific norms should be used when selecting IAPS pictures.

Due to the paucity of studies providing categorization of IAPS pictures in a British sample, in the current adult study, pictures were selected from the Mikels et al. (2005) paper that had been rated as representing sadness and fear only, not mixed emotions, and only those that in their sample did not show gender differences in their valence and arousal ratings. Due to the small sample size in both our studies, sex differences and cross-cultural differences in ratings were not taken into account.

The nature of work for this research is rooted in empirical software engineering using a controlled experiment method. The system, which will recognize the participants' emotions and control the materials delivery, is the independent variable that will be manipulated to measure its effect on the dependent variable, which will be the participants' emotional state during the assessment.

3.1. Participants

For the adult-related experiment, nineteen participants were recruited, the demographics of whom are depicted in **Table 1**. Participants were a combination of undergraduate psychology students from the University of Westminster, who were awarded 1 h of research participation credit, and colleagues and acquaintances recruited by word of mouth. For the child-related experiment, eleven children were recruited (five females, six males, with a mean age of 11.5 years, SD 3.24, and an age range of 7–16). They were recruited by opportunistic sampling and word of mouth. Written parental consent and verbal child consent was obtained.

TABLE 1 | Mean age and standard deviation for 19 participants.

	Males (*n* = 8)	**Females (*n* = 11)**
Age, years, mean ± SD	33.10 ± 16.06	28.55 ±10.48
Age range in years	19–61	19–46

3.2. Ethics and Regulatory Framework

This study was carried out in accordance with the Ethics Code of the University of Westminster. This includes the assurance that data about an individual will be held securely, handled in accordance with the Data Protection Act 1998, and disposed of in line with Westminster's retention policy. The Ethics Code[1] and the Data Protection Policy[2] are available from the University of Westminster's Website. For the child-related experiment, written parental consent and verbal child consent was obtained.

In order to protect research participants and research staff involved with unpleasant IAPS protocols, specific measures were taken to ensure that the risk of negative psychological consequences was minimized for all parties involved, and the nature of the images shown was fully explained to both research staff and participants. As the researchers were interested in emotions that may be evoked in a hospital waiting room environment, no images involving mutilations or sexually arousing images were included, and the unpleasant images selected for the study had been identified as more likely to have been rated as evoking discrete emotions such as fear and sadness.

The researchers observed each participant during the study, and there were procedures in place so that if it was noticed that a participant was becoming distressed or emotionally upset, the researchers would ask the participant if they needed anything or would like to take a break. If a participant was distressed, the researcher would also get in contact with them later in the day to confirm that they were feeling less distress and that they had sought any help they may have needed. In addition, participants were provided with details of the University Counseling Services and, following completion of the study, all participants were fully debriefed.

3.2.1. Data

Participants are referred to using unique numbers, and no collected data contains participant-identifying information. The consent form requires a signature and/or the initials of the participant, but this is kept separately from any data and is held in a secure file that is only available to the research team. Data are stored on a password-protected computer on University premises, and only the research team have access. Data stored on the laptop and external hard-drive are encrypted. Only members

[1]https://www.westminster.ac.uk/sites/default/public-files/general-documents/research-code-of-good-practice.pdf

[2]https://www.westminster.ac.uk/about-us/our-university/corporate-information/information-compliance-records-management-and-information-security/personal-data-protection/

of the research team have access to the key. Participants had been made aware that their facial expressions will be photographed and videoed. No photographs or videos will be published without the participant's explicit consent.

3.3. Materials

3.3.1. Photographic Stimuli

The International Affective Picture System (IAPS) (Lang et al., 1997) is a set of more than 900 standardized pictures that has been widely used in the study of emotion and attention, with more than 2700 citations. When used with tools for the collection of subjective affective ratings such as the Self-Assessment Manikin (SAM) (Lang et al., 1997) or the Affective Slider (Betella and Verschure, 2016), insights into the dimensional aspects of emotion are derived. The set of pictures includes pictures such as snakes, accidents, kittens, babies, and everyday items such as chairs. Based on previous emotional ratings, 80 pictures were selected for the current study, 50% neutral, 10% pleasant, and 40% unpleasant. The pictures selected were based on a study by Mikels et al. (2005) who attempted to provide categories for the IAPS pictures based on the ratings of pictures by a sample of 120 participants in the United States. Although it can be difficult to elicit emotional responses in a laboratory environment, particularly discrete emotions, the pictures in the current study were selected to try to evoke emotions in a hospital waiting room environment, so pictures that were rated as likely to represent sadness and fear were chosen as well as those rated for happiness. Pictures in the dataset that had been rated as showing a mixture of emotions, e.g., fear, anger, and anxiety were excluded as were pictures that had shown sex differences in their ratings.

Normative ratings are available for the pictures based on a nine-point scale. Mikels et al. (2005) selected the images based on minimum criteria. The negative images met the minimum criterion that they be less than the neutral midpoint of 5 (mean pleasure rating = 3.05, *SD* = 0.84, and mean arousal rating = 5.56, *SD* = 0.92). The positive subset were selected as positively valenced on the criterion of being equal to or greater than 5 (mean pleasure rating = 7.05, *SD* = 0.63, mean arousal rating = 4.87, *SD* = 0.98). As we were particularly interested in negative emotions, the images selected were based on emotions that may be experienced in a hospital waiting room, such as fear and sadness. Images of a sexual nature or ones that evoked disgust were not relevant to the current study.

3.3.2. Affective Digital Slider

The affective digital slider is a tool that provides a self-assessment scale for the measurement of human emotions that does not require written instructions. There are two sliders, one measuring arousal, ranging from calm to excited, and one measuring affective valence, ranging from pleasant to unpleasant (see **Figure 1**). Each slider measures a single value on a continuous normalized scale ranging from 1 to 9 with a central value equal to 5 and a minimum resolution 0.01; 9 represents a high rating on each dimension (i.e., high pleasure, high arousal) and 1 represents a low rating on each dimension (i.e., low pleasure, low arousal).

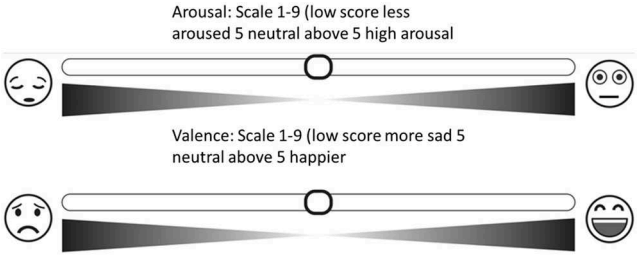

FIGURE 1 | Overview of the study protocol including the "Affective Slider" (Betella and Verschure, 2016) (AS), which measures arousal (top) and pleasure (bottom) on a continuous scale.

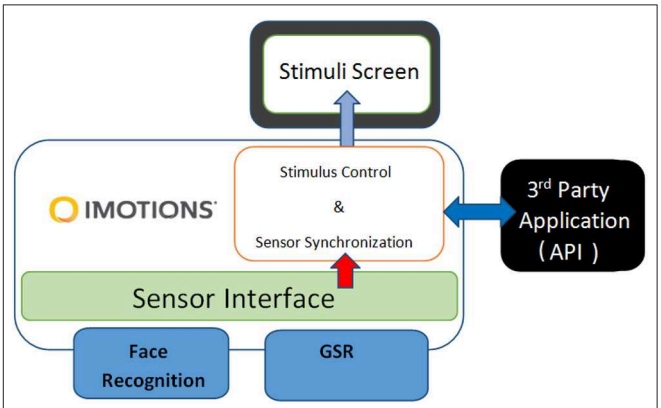

FIGURE 2 | System diagram of the multimodal human behavior study.

3.3.3. Subjective Data

In addition to providing dimensional data using the digital slider, participants were asked to select from a list the word that best described the predominant emotion that they felt after viewing each picture. The list included the words happy, sad, fear, neutral, and disgust; there was also an option for the participants to select 'other' and write their own description. For data analysis purposes, these subjective ratings were then coded as negative, neutral, or positive (such that happy was rated as positive, sad, fear were negative).

3.4. Biometric Research Platform

Galvanic Skin Response (GSR) and facial expressions were measured using the iMotions Biometric Research Platform 6.0, iMotions A/S, Copenhagen, Denmark, 2016. An Application Programming Interface (API) module was used together with the iMotions platform and the Galvanic sensor to monitor and control in real time the connections with the biometric sensors through TCP ports, and the data flow of the experiment, with time, sequence number, and stimulus name and type assigned to variables (see **Figure 2**). Facial expressions were recorded for analysis via a webcam (Logitech HD Pro Webcam C920).

iMotions can detect changes in key face features such as brows, eyes, and lips and analyze the basic emotions of the recorded face. Researchers can choose between two different algorithms to

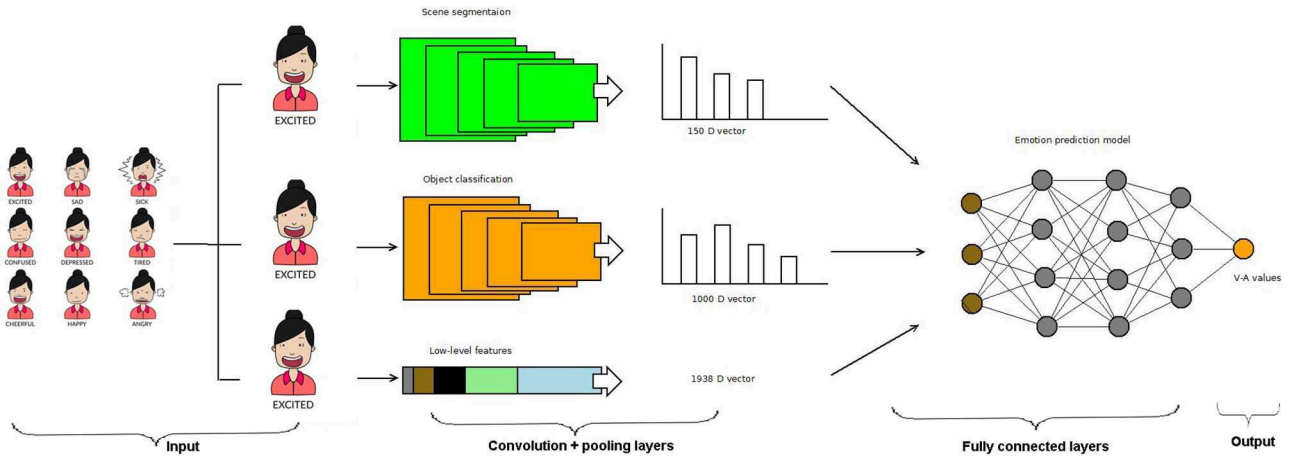

FIGURE 3 | The general pipeline of the iMotions recognition system. Adapted from Kim et al. (2018). Female Mood Avatars by Namnso Ukpanah[3].

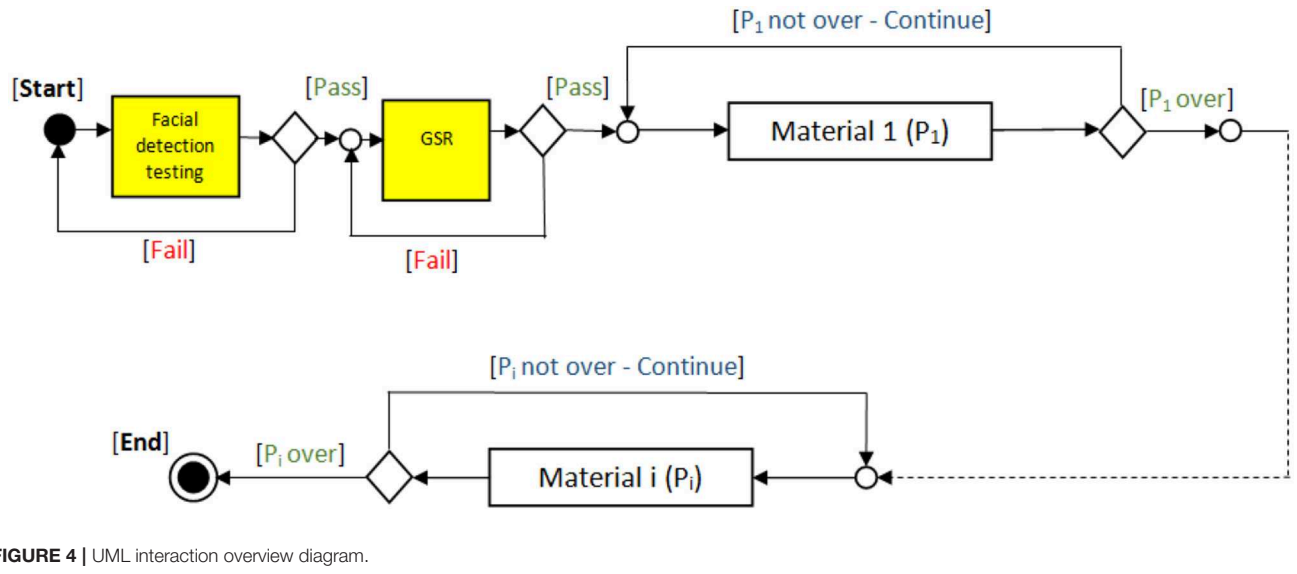

FIGURE 4 | UML interaction overview diagram.

classify emotions from facial expressions in iMotions's platform: the FACET module, based on the FACET algorithm (Littlewort et al., 2011), and the AFFDEX module, based on the AFFDEX algorithm by Affectiva Inc. (El Kaliouby and Robinson, 2005; McDuff et al., 2010). Affectiva is an API for emotion recognition using deep learning. It is said to have nearly 6 million faces as an emotion database in order to provide great accuracy[4]. These algorithms detect facial landmarks and apply a set of rules based on psychological theories and statistical procedures to classify emotions (Li and Deng, 2018; Stöckli et al., 2018). Different algorithms, like AFFDEX and FACET, use distinct statistical procedures, facial databases, and facial landmarks to train the machine learning procedures and ultimately classify emotions (Kim et al., 2018). For all our experiments, we have used the AFFDEX algorithm. iMotions classifies the seven basic emotions (joy, anger, surprise, fear, contempt, sadness, and disgust) and provides a confidence rating for the probability that an emotion is being expressed. For data analysis purposes, joy was coded as positive; anger, fear, contempt, sadness, and disgust were coded as negative; and data reflecting surprise was excluded, as surprise can be either positive or negative in valence. **Figure 3** shows the iMotions architecture diagram with an image or video sequence as input, the feature maps based on convolution and pooling layers, the fully connected layers, and the output, which can be happiness, sadness, or any of the seven emotions classified by iMotions.

Galvanic skin response was measured from the phalanx of the index and middle finger of the nondominant hand using 1

[3]https://blog.affectiva.com/the-worlds-largest-emotion-database-5.3-million-faces-and-counting

[4]https://search.creativecommons.org/photos/21d1d4c8-a166-4ac2-9100-f6e33ab7b316

cm^2 Ag/AgCl (silver/silverchloride) electrodes placed in reusable snap-on Velcro straps. For each participant, the GSR recorded in microvolts (μV) was segmented into 8-s intervals for each picture, and the mean of each participant's 2-min baseline measure was subtracted from the peak of each segment.

The API module was designed to receive the biometric sensor data, analyse it, and control the delivery of the presented materials. The Unified Modeling Language (UML) interaction overview diagram (see **Figure 4**) shows how the system starts by testing the biosensors, starting with the camera for facial detection, then the GSR sensor. After passing the tests successfully, the participant engages with the first material (P1), while the API reads and analyses the data provided by the sensors. The material will continue playing to the end before moving to the next. This process will continue in the same pattern through the rest of the materials.

Figure 5 presents a flow chart for the pilot study of the automated emotion assessment in adults and children. The API continuously reads/monitors the data and provides control signals accordingly until the emotion assessment session is completed.

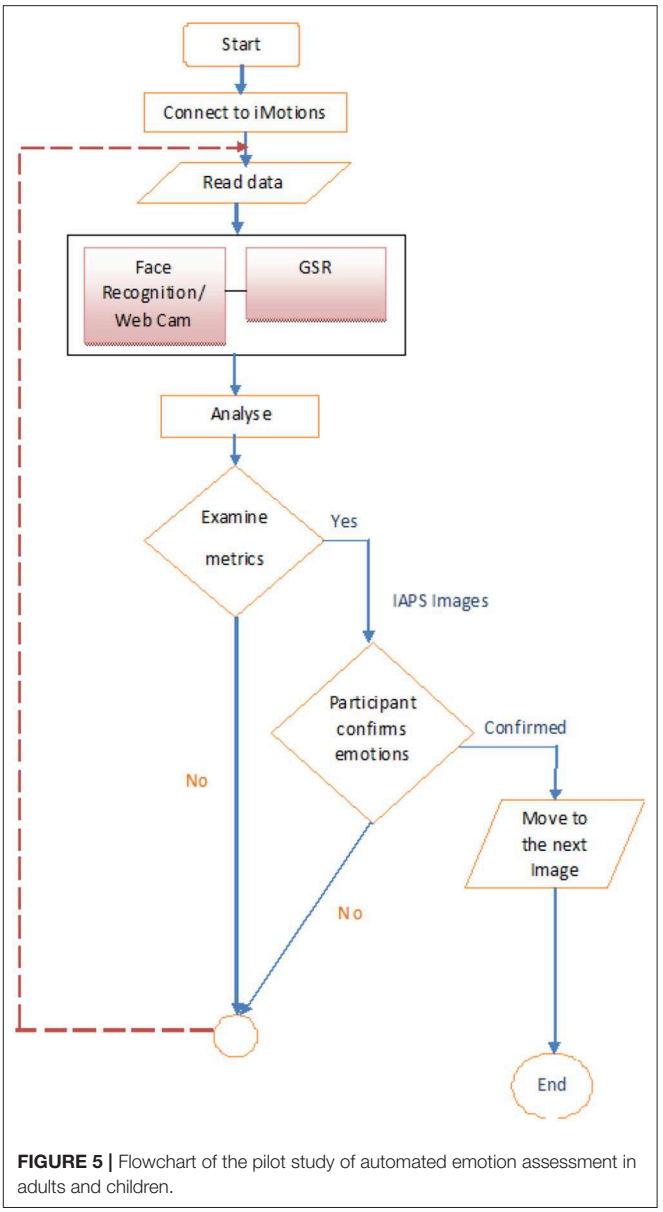

FIGURE 5 | Flowchart of the pilot study of automated emotion assessment in adults and children.

3.5. Protocol

Participants were seated comfortably in front of a computer screen. They were advised that their facial expressions in response to each photograph would be recorded via a webcam and recorded with iMotions facial expression analysis software. Participants were recorded individually in a quiet laboratory at the university and were asked to rate a set of 80 photographs selected from IAPS. The order of the photographs was pseudo-random such that each emotive photograph was preceded by a neutral photograph. Each photograph was presented on a computer screen for 8 s. Participants were instructed to maintain their attention on the screen for the whole time that the image was present. Galvanic skin response was recorded from each participant at baseline for 2 min, during which time the participants were asked to relax and close their eyes. GSR was then recorded throughout the study. Standardized instructions were read to each participant based on the Self-Assessment Manikin instructions (Bradley and Lang, 1994) amended for use with the Affective Slider (Betella and Verschure, 2016) with the instruction to "move the sliders to express how you actually felt while watching the picture." Participants were asked to view and rate four photographs that were similar to those used in the study to familiarize themselves with the rating scales and to ensure that they were happy to participate. Written consent forms were completed.

Figure 6 shows the experimental set up with the participant. Following each photograph, the screen showed the rating page, and the participant was asked to provide their subjective affective ratings for pleasure and arousal using the Affective Slider and then to select the word that indicated the predominant emotion that they felt when viewing the picture. Participants were advised that if the word that described their emotion was not there, they should select 'other' and type the word that best describes how they felt in response to the photograph in the space provided. The rating screen stayed in place until ratings had been made. A further screen then appeared for 4 s, advising the participant to prepare to view the next slide.

3.5.1. Additional Notes for the Child-Related Experiment

Each child was rewarded for their participation with a £10 "Love to Shop" voucher. They were advised that they could withdraw their participation at any point without penalty. Recording of the data took place in the participants' home environment. Thirty-two images (16 neutral, eight positive and eight negative) were selected from IAPS for the study with children based on those used in a study by McManis et al. (2001), where they had been judged by teachers to be appropriate for viewing by children in the age range 7–14 years. The images covered a wide range of affective content, and each emotive image was preceded by

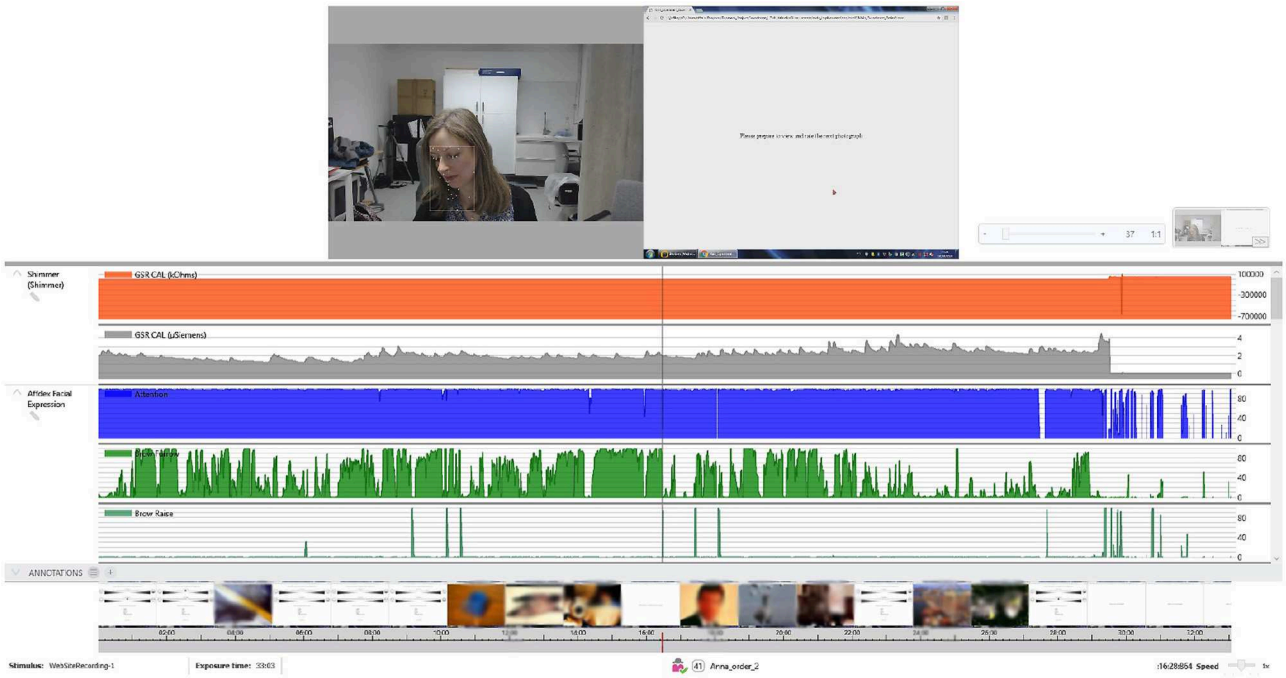

FIGURE 6 | Overview of the study protocol. The top row shows the participant together with the facial landmark points assigned by the iMotions facial recognition algorithm and a screen with instructions. The middle row shows the GSR and the Affdex Facial expression metrics. The bottom row shows in a timeline the response of the participant per image, and the Affective Slider, which, as can be seen in **Figure 1**, measures arousal (top) and pleasure (bottom) on a continuous scale. Due to copyright issues associated with the IAPS images, all images at the bottom have been blurred.

a neutral image. Standardized instructions based on the SAM for children, adjusted for use with the Affective Digital Slider, were read to each participant. Participants were seated in front of a laptop with a built-in camera that recorded their facial expressions. Parents were allowed to be in the room if they requested to be.

4. STATISTICAL ANALYSIS AND RESULTS

The following section provides details of the treatment of results from both experiments. The participant's valence, arousal, and GSR scores are subjected to Analysis of Variance (ANOVA), whilst participants' subjective ratings and iMotions classification are subjected to analysis using Chi-square.

4.1. Study With Adults

Participants' responses on the digital slider scales for valence and arousal were recorded for each picture, as was GSR response. Mean scores and standard deviations (SD) of participants' ratings (on a scale of 1-9) for each of the IAPS negative, neutral, and positive pictures and their GSR response recorded in microvolts (μV) are shown in **Table 2**. GSR data for one participant was excluded due to technical difficulties during the recording.

Low scores (i.e., $<$ 5) indicate negative valence and low arousal, scores around 5 indicate neutral valence and little or no reported arousal, and higher scores (i.e., $>$5) indicate positive valence and high arousal. The descriptive statistics reported in

TABLE 2 | Mean (+SD) valence and arousal ratings for the Affective Digital Slider ($n = 19$) and GSR data ($n = 18$) for each IAPS picture category.

	IAPs category		
Measure	**Negative ($n = 30$)**	**Neutral ($n = 40$)**	**Positive ($n = 10$)**
Valence	3.49 (±0.91)	5.02 (±0.23)	6.38(±1.19)
Arousal	5.78 (±1.13)	4.55 (±1.03)	4.16 (±1.00)
GSR(μV)	9.16 (±27.67)	8.78 (±27.50)	11.73 (±28.02)

Table 2 suggest that, for valence, as expected, negative pictures have the lowest mean score, suggesting that they evoked negative emotions, neutral pictures have a mean score at the mid-point of the scale, suggesting that neither positive or negative emotion was evoked, and positive pictures have the highest mean score, suggesting that more positive emotions were reported.

In terms of arousal, it was expected that both positive and negative images would have a higher mean arousal rating than neutral images. However, the descriptive statistics suggest that negative images elicited a higher mean arousal rating, while the ratings for neutral and positive images were similar. In addition, mean GSR scores (μV) indicate that positive pictures and negative images elicited a greater GSR response than neutral pictures, with positive pictures eliciting the greatest GSR.

In order to establish whether any of these differences were significant, repeated-measures one-way ANOVA was conducted. The within-subjects factor, the type of IAPS image, had three

levels: negative, neutral, and positive. There was a significant effect of the type of IAPS image on the valence score [$F_{(2,36)} = 39.989, p < 0.01$]. Bonferroni-corrected simple effects reveal that, using the digital slider, positive images were rated as significantly more positive than both neutral and negative images ($t = 4.871, df = 18, p < 0.05$; $t = 6.538, df = 18, p < 0.05$) and negative images were rated as significantly more negative than neutral images ($t = 1.420, df = 18, p = 0.173$).

There was a significant effect of the type of IAPS image on the arousal score [$F_{(2,36)} = 10.500, p < 0.01$]. Bonferroni-corrected simple effects reveal that, using the digital slider, negative images were rated as significantly more arousing than positive images ($t = 3.872, df = 18, p = 0.001$) but not significantly more arousing than neutral images ($t = 3.093, df = 18, p = 0.006$), and there was not a significant difference in arousal rating between neutral and positive images ($t = 7.636, df = 18, p < 0.05$).

There was a statistically significant effect of the type of IAPS image on the GSR score [$F_{(2,34)} = 25.037, p < 0.001$]. Bonferroni simple effects reveal that positive images provoked a significantly higher GSR than neutral or negative images ($t = 5.387, df = 17, p = 0.001$; $t = 4.758, df = 17, p < 0.001$); there was no significant difference in GSR between neutral and negative images ($t = 2.341, df = 17, p = 0.032$). These findings suggest that, when using the digital sliders, participants generally showed agreement with the IAPS classification of the images in terms of their valence rating and that negative images were more arousing when rated using the digital slider but that positive images elicited the greatest GSR.

As ratings using the affective digital slider show that participants generally rate the images in accordance with the IAPS classification and neutral images were only included in the study to bring participants' valence and arousal back to neutral between each of the emotive images, analysis of the association of subjective ratings and picture type and iMotions data and picture type focused on the negative and positive images.

A Chi-squared test was conducted to test for an association between the type of image displayed (negative or positive) and subjective rating of participants, who selected a word that best described how each picture made them feel (negative or positive). Results show a significant association between IAPS picture type and subjective rating ($X^2 = 192.700, df = 1, p < 0.001$). Participants rated negative pictures as negative in 85.6% of cases and rated positive pictures as positive in 75.9% of cases (see **Table 3** for observed and expected counts). The results show that participants were more likely to rate a negative picture as negative and a positive picture as positive.

The results of a Chi-squared test for an association between the type of picture displayed and the classification of the facial expression by iMotions software show a significant relationship between IAPS picture type and iMotions ($X^2 = 32.233, df = 1, p < 0.001$). The iMotions software identified negative facial expressions in participants viewing negative images in 95.8% of cases and identified positive facial expressions in participants viewing positive images in 23% of cases. Interestingly, iMotions also identified negative emotional reactions to 77% of the positive pictures (see **Table 4** for expected and observed counts).

TABLE 3 | Participants' ratings (negative, positive) of negative and positive images.

		IAPS picture type		
Subjective rating	**Count**	**Negative**	**Positive**	**Row totals**
Negative	Observed	357 (85.6%)	35 (24.1%)	392
	Expected	290.9	101.1	
Positive	Observed	60 (14.4%)	110 (75.9%)	170
	Expected	126.1	43.9	
Columns totals		417	145	562 (grand total)

TABLE 4 | iMotions classification (negative, positive) of participants' facial expressions to negative and positive images.

		IAPS picture type		
iMotions identifier	**Count**	**Negative**	**Positive**	**Row totals**
Negative	Observed	277 (95.8%)	77 (77%)	354
	Expected	263.0	91.0	
Positive	Observed	12 (4.2%)	23 (23%)	35
	Expected	26.0	9.0	
Columns totals		289	100	389 (grand total)

TABLE 5 | Mean (+SD) valence and arousal ratings for the Affective Digital Slider ($n = 11$) and GSR data for each IAPS picture category.

	IAPs category		
Measure	**Negative ($n = 8$)**	**Neutral ($n = 16$)**	**Positive ($n = 8$)**
Valence	5.26 (±0.49)	5.01 (±0.42)	5.40 (±0.67)
Arousal	5.62 (±0.67)	4.43 (±1.13)	5.87 (±0.87)
GSR(μV)	3994.85 (±203.60)	3997.65 (±204.59)	3999.27 (±204.16)

The results show that iMotions software was more likely to identify negative facial expressions in response to negative images but interestingly, more negative facial expressions to positive pictures.

4.2. Study With Children

Participants' responses on the digital slider scales for valence and arousal were recorded for each picture, as was GSR response. Mean scores and standard deviations (SD) of participants' ratings (on a scale of 1-9) for each of the IAPS negative, neutral, and positive pictures and their GSR response recorded in microvolts (μV) are shown in **Table 5**.

The descriptive statistics reported in **Table 5** suggest that, for valence, the pictures did not evoke reportable negative or positive emotions. In terms of arousal, both positive and negative images have a higher mean arousal rating than neutral images. In addition, mean GSR scores (μV) indicate that positive pictures elicited a greater GSR response than both neutral and negative pictures. To establish whether any of these differences were significant, repeated-measures one-way ANOVA was conducted.

The within-subjects factor, the type of IAPS image, had three levels, negative, neutral and positive. There was no significant

TABLE 6 | Participants' ratings (negative, positive) of negative and positive images.

		IAPS picture type		
Subjective rating	**Count**	**Negative**	**Positive**	**Row totals**
Negative	Observed	33 (50%)	40 (50.6%)	73
	Expected	33.2	39.8	
Positive	Observed	33 (50%)	39 (49.4%)	72
	Expected	126.1	43.9	
Columns totals		66	79	145 (grand total)

TABLE 7 | iMotions classification (negative, positive) of participants' facial expressions to negative and positive images.

		IAPS picture type		
iMotions Identifier	**Count**	**Negative**	**Positive**	**Row totals**
Negative	Observed	25 (71.4%)	24 (53.3%)	49
	Expected	21.4.0	27.6	
Positive	Observed	10 (28.6%)	21 (46.7%)	31
	Expected	13.6	17.4	
Columns totals		35	45	80 (grand total)

effect of the type of IAPS image on the valence score [$F_{(2,20)} = 1.744, p = 0.200$]. This indicates that the pictures did not evoke reported emotional responses in the children. There was a significant effect of the type of IAPS image on the arousal score [$F_{(2,20)} = 6.028, p < 0.01$]. Bonferroni-corrected simple effects (p should be $\leq$0.0016 to reach significance), however, did not reach significance when comparing negative with neutral images ($t = 2.390, df = 10, p = 0.038$) and positive with neutral images ($t = 2.654, df = 10, p = 0.006$).

There was a statistically significant effect of the type of IAPS image on the GSR score [$F_{(2,20)} = 22.193, p < 0.001$]. Bonferroni-corrected simple effects reveal that positive images had a significantly higher GSR than negative images ($t = 5.442, df = 10, p < 0.001$), but there was no significant difference in GSR between neutral and negative images ($t = 3.804, df = 10, p < 0.03$) or between neutral and positive images; ($t = 4.758, df = 17, p < 0.002$). These findings suggest that, when using the digital sliders, participants did not report either negative or positive responses or higher arousal to negative and positive images, as would be expected. GSR, however, was higher in response to positive images.

In order to explore whether subjective ratings and iMotions measures of facial expression were associated with the type of picture presented, Chi-squared tests were conducted. A Chi-squared test was conducted to test for an association between the type of image displayed (negative or positive) and subjective rating of participants, who selected a word that best described how each picture made them feel (negative or positive). The results show no association between IAPS picture type and subjective ratings ($X^2 = 0.006, df = 1, p = 1$). Participants rated negative pictures as negative in 50% of cases and rated positive pictures as positive in 49.4% of cases (see **Table 6** for observed and expected counts). The results show that participants were just as likely to rate a negative picture as positive and a positive picture as negative.

The results of a Chi-squared test for an association between the type of picture displayed and the classification of the facial expressions by iMotions software did not show an association between IAPS picture type and iMotions ($X^2 = 2.716, df = 1, p = 0.112$). The iMotions software identified negative facial expressions in participants viewing negative images in 71.4% of cases where a negative picture was shown and identified positive facial expressions in participants viewing positive images in 46.7% of cases (see **Table 7** for expected and observed counts). The results show that iMotions software was more likely to identify negative facial expressions in response to negative images but also to positive pictures.

5. DISCUSSION

The findings in the current study are interesting for a number of reasons. Rating the IAPS images using the Affective Digital Slider produced differing results for adults and children. Adults rated positive images with a higher valence, which was representative of the images making them feel happier than neutral images and negative images with lower valence, suggesting that the negative images made them feel sad/fearful. This is what would be expected and would appear to corroborate the IAPS classification of the images, as has been shown in many previous studies. However, it should be noted that the mean ratings for the positive images, although consistently higher than neutral, were not that much higher. In contrast, the children did not rate the negative, neutral, or positive images as differing in valence or arousal. These findings are in alignment with those of Vetella and Verscure (2016), who also found that standardized sets of stimuli, such as IAPS, may not be as effective as they once were at evoking emotions, due to our exposure to highly arousing stimuli in the media and general desensitization. In addition, the respective associations of subjective ratings and iMotions classification of facial expressions with picture category suggest that, for adults, subjective rating is better at identifying emotions than biometric software. It is unsurprising that participants show higher reliability than iMotions in classifying positive and negative images, given the importance of facial expression to human communication. In children, subjective ratings and the iMotions identification of expressions were at the level of chance. An additional factor that must be considered is that researchers in both studies noted little change in the facial expression of participants as they observed the IAPS images, once again suggesting that the stimuli may not have sufficiently evoked the emotions of the participants.

Other explanations for the difference in findings between children and adults could be due to the IAPS pictures selected and the environment in which the studies took place. Children were shown different and fewer images to ensure that the images were age-appropriate and that the task was not too onerous. The adult experiments took place in a laboratory at the University, whilst the children performed the task in their home environment. In the home environment, conditions such as lighting could not be

controlled in the same way as in the laboratory, and this may have led to the biometric platform not detecting all facial movements. One of the difficulties with evoking and categorizing emotions in a laboratory scenario and not only through the use of pictures is establishing how each picture will actually make the individual feel—this will vary between participants. What may evoke fear in one participant may evoke anger, or indeed a mixture of more than one emotion, in another. In future studies, it is important to use more provocative stimuli (for example, emotive video clips) to ensure that emotions are sufficiently elicited and a larger sample and to ensure maximal environmental conditions for the use of the biometric platform.

While the biometrics facial recognition industry has grown, facial movements and expressions may not always be a reliable indication of how someone is feeling. Studies have shown that humans make assessments about other people's emotions based on factors including body language and tone of voice. As such, many emotion detection algorithms that have been developed in the last two decades are still facing problems with accuracy, complexity, and real-world implementation due to the irregularities in the complexity of models and unpredictability between expression categories. These approaches should thus always be used responsibly, especially when used in crime-detection applications.

6. CONCLUSION

In this paper, we contributed to the lively debate about and criticism of the effectiveness of automated emotion recognition approaches by attempting to question the following aspects: (a) the adequacy of well-established techniques such as the International Affective Picture System (IAPS), (b) the adequacy of state-of-the-art biometric research platforms, (c) the extent to which emotional responses may be different between children or adults. Our initial statistical analysis and results indicate that although there is, in general, an alignment between expected (IAPS) and observed (iMotions) responses for negative images, there is an interesting discrepancy in the expected and observed responses for positive images. This may be for many reasons ranging from incorrect classification of images in IAPS to incorrect classifications of responses by the biometric research system, iMotions, to significant changes in the emotional responses of the human population. In the future, we plan to dig deeper into the correlations among all of the significant variables by addressing all aspects of the experiment: facial recognition, subjective classification of responses, collection of images (IAPS), and different groups, i.e., adults and children. This multi-dimensional, multi-variate analysis will help shed more light on the real causes of such problems with automated emotion recognition, as well as into the limitations of current state-of-the-art approaches and technologies.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

This study was approved by the Psychology Ethics Committee at the University of Westminster. The participants provided their written informed consent to participate in this study. For those under the age of 16, written informed parental consent and verbal child consent was obtained. Written informed consent was obtained from the individual(s) AND/OR minor(s)' legal guardian/next of kin for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

All authors contributed equally for the experimental analysis and preparation of the manuscript.

FUNDING

This research was funded by the University of Westminster through a Phase 4 - Excellence with Impact 2016 grant under project code RH411310.

ACKNOWLEDGMENTS

We would also like to thank Dr. Nina Smyth for assistance with statistical analysis.

REFERENCES

Acharya, U., Oh, S., Hagiwara, Y., Tan, J., Adeli, H., and Subha, D. (2018). Automated EEG-based screening of depression using deep convolutional neural network. *Comput. Methods Progr. Biomed.* 161, 103–113. doi: 10.1016/j.cmpb.2018.04.012

Barke, A., Stahl, J., and Kröner-Herwig, B. (2012). Identifying a subset of fear-evoking pictures from the iaps on the basis of dimensional and categorical ratings for a german sample. *J. Behav. Ther. Exp. Psychiatry* 43, 565–572. doi: 10.1016/j.jbtep.2011.07.006

Betella, A., and Verschure, P. (2016). The affective slider: a digital self-assessment scale for the measurement of human emotions. *PLoS ONE* 11:e0148037. doi: 10.1371/journal.pone.0148037

Bonanno, G., and Keltner, D. (2004). Brief report: The coherence of emotion systems: Comparing "on-line" measures of appraisal and facial expressions, and self-report. *Cogn. Emot.* 18, 431–444. doi: 10.1080/02699930341000149

Bosse, T., Gratch, J., Hoorn, J. F., Portier, M., and Siddiqui, G. F. (2010). Comparing three computational models of affect. *Adv. Pract. Appl. Agents Multiagent Syst.* 70, 175–184. doi: 10.1007/978-3-642-12384-9_22

Boucsein, W., Fowles, D., Grimnes, S., Ben-Shakhar, G., Roth, W., Dawson, M., et al. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology* 49, 1017–1034. doi: 10.1111/j.1469-8986.2012.01384.x

Bradley, M., and Lang, P. (1994). Measuring emotion - the self-assessment mannequin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* 25, 49–59. doi: 10.1016/0005-7916(94)90063-9

Colombetti, G. (1995). From affect programs to dynamical discrete emotions. *Philos. Psychol.* 22, 407–425. doi: 10.1080/09515080903153600

Dalgleish, T., and Power, M. (eds.). (1999). *Handbook of Cognition and Emotion*. Chichester, UK: Wiley Online Library.

Ekman, P. (1993). Facial expression and emotion. *Amer. Psychol.* 48, 384–392. doi: 10.1037/0003-066X.48.4.384

Ekman, P. (2003). *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life*. New York, NY: Times.

Ekman, P. (2016). What scientists who study emotion agree about. *Perspect. Psychol. Sci.* 11, 31–34. doi: 10.1177/1745691615596992

Ekman, P., and Friesen, W. V. (1978). *Facial Action Coding System (FACS). A Technique for the Measurement of Facial Actions*. Palo Alto, CA: Consulting Psychologist Press.

El Kaliouby, R., and Robinson, P. (2005). "Real-time inference of complex mental states from facial expressions and head gestures," in *Real-Time Vision for Human-Computer Interaction* (Boston, MA), 181–200.

Feidakis, M., Daradoumis, T., and Caballe, S. (2011). "Emotion measurement in intelligent tutoring systems: what, when and how to measure," in *Third International Conference on Intelligent Networking and Collaborative Systems* (Fukuoka: IEEE).

Fernández-Rodríguez, A., Velasco-Alvarez, F., Medina-Julía, M., and Ron-Angevin, R. (2019). Evaluation of emotional and neutral pictures as flashing stimuli using a p300 brain-computer interface speller. *Neural Eng.* 16, 1–12. doi: 10.1088/1741-2552/ab386d

Fontaine, J. R., Scherer, K. R., Roesch, E. B., and Ellsworth, P. (2007). The world of emotion is not two-dimensional. *Psychol. Sci.* 18, 1050–1057. doi: 10.1111/j.1467-9280.2007.02024.x

George, J. M. (1996). "Trait and state affect," in *Individual Differences and Behavior in Organizations*, ed K. R. Murphy (San Francisco, CA: Jossey-Bass), 145–171.

Greenwald, M., Cook, E., and Lang, P. (1989). Affective judgment and psychophysiological response: dimensional covariation in the evaluation of pictorial stimuli. *J. Psychophysiol.* 3, 51–64.

Guan, S., Zhao, K., and S., Y. (2019). Motor imagery EEG classification based on decision tree framework and riemannian geometry. *Comput. Intell. Neurosci.* 2019:5627156. doi: 10.1155/2019/5627156

Hajian, S., and Bonchi, F. (2016). "Algorithmic bias: from discrimination discovery to fairness-aware data mining," in *KDD 2016 Tutorial*. San Francisco, CA.

Hou, Y., and Chen, S. (2019). Distinguishing different emotions evoked by music via electroencephalographic signals. *Comput. Intell. Neurosci.* 2019:3191903. doi: 10.1155/2019/3191903

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.* 1–1. doi: 10.1109/TAFFC.2019.2901456

Jahmunah, V., Oh, S., Rajinikanth, V., Ciaccio, E., Cheong, K., Arunkumar, N., et al. (2019). Automated detection of schizophrenia using nonlinear signal processing methods. *Artif. Intell. Med.* 100:101698. doi: 10.1016/j.artmed.2019.07.006

Kim, H. -R., Kim, Y. -S., Kim, S. J., and Lee, I. -K. (2018). Building emotional machines: recognizing image emotions through deep neural networks. *IEEE Trans. Multimedia* 20, 2980–2992. doi: 10.1109/TMM.2018.2827782

Kunz, M., Seuss, D., Hassan, T., Garbas, J., Siebers, M., Schmid, U., et al. (2017). Problems of video-based pain detection in patients with dementia: a road map to an interdisciplinary solution. *BMC Geriatr.* 17:33. doi: 10.1186/s12877-017-0427-2

Lang, P. (1969). "The mechanics of desensitization and the laboratory study of human fear," in *Behavior Therapy: Appraisal and Status* (New York, NY: McGraw-Hill), 160–191.

Lang, P., Bradley, M., and Cuthbert, B. (1997). "International affective picture system (IAPS): technical manual and affective ratings," in *NIMH Center for the Study of Emotion and Attention* (Florida, FL), 39–58.

Leo, M., Del Coco, M., Carcagni, P., Distante, C., Bernava, M., Pioggia, G., et al. (2015). "Automatic emotion recognition in robot-children interaction for asd treatment," in *Proceedings of the IEEE International Conference on Computer Vision Workshops* (Santiago), 145–153.

Li, S., and Deng, W. (2018). Deep facial expression recognition: a survey. *CoRR* abs/1804.08348.

Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., et al. (2011). "The computer expression recognition toolbox (CERT)," in *Proceedings from 2011 I.E. International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011)* (Santa Barbara, CA), 298–305.

McDuff, D., El Kaliouby, R., Kassam, K., and Picard, R. (2010). "Affect valence inference from facial action unit spectrograms," in *Proceedings from 2010 I.E. Computer Society Conference on Computer Vision and Pattern Recognition – Workshops* (San Francisco, CA), 17–24.

McManis, M., Bradley, M., Berg, W., Cuthbert, B., and Lang, P. (2001). Emotional reactions in children: verbal, physiological, and behavioral responses to affective pictures. *Psychophysiology* 38, 222–231. doi: 10.1111/1469-8986.3820222

Mikels, J. A., Fredrickson, B. L., Larkin, G. R., Lindberg, C. M., Maglio, S. J., and Reuter-Lorenz, P. A. (2005). Emotional category data on images from the international affective picture system. *Behav. Res. Methods* 37, 626–630. doi: 10.3758/BF03192732

Ortony, A., Clore, G., and Collins, A. (1987). *The Cognitive Structure of Emotions*. Cambridge: Cambridge University Press.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. *Emotion Theory Res. Exp.* 1, 3–33. doi: 10.1016/B978-0-12-558701-3.50007-7

Plutchik, R. (2003). *Emotions and Life*. Washington, DC: American Psychological Association.

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

Prinz, J. (2004). *Gut Reactions: A Perceptual Theory of Emotion*. New York, NY: Oxford University Press.

Reisenzein, R., Studtmann, M., and Horstmann, G. (2013). Coherence between emotion and facial expression: evidence from laboratory experiments. *Emot. Rev.* 5, 16–23. doi: 10.1177/1754073912457228

Rubin, D. C., and Talerico, J. M. (2009). A comparison of dimensional models of emotion. *Memory* 17, 802–808. doi: 10.1080/09658210903130764

Russell, J. A. (1980). A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 1161–1178. doi: 10.1037/h0077714

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychol. Rev.* 110, 145–172. doi: 10.1037/0033-295X.110.1.145

Shen, L., Wang, M., and Shen, R. (2009). Affective e-learning: using "emotional" data to improve learning in pervasive learning environment. *J. Educ. Technol. Soc.* 12, 176–189.

Sikka, K., Ahmed, A., Diaz, D., Goodwin, M. S., Craig, K., Bartlett, M., et al. (2015). Automated assessment of children postoperative pain using computer vision. *Pediatrics* 136, 124–131. doi: 10.1542/peds.2015-0029

Stöckli, S., Schulte-Mecklenbeck, M., and Borer, S. E. A. (2018). Facial expression analysis with affdex and facet: a validation study. *Behav. Res. Methods* 50, 1446–1460. doi: 10.3758/s13428-017-0996-1

Suzan, A., and Mariofanna, M. (2016). Real time face expression recognition of children with autism. *Int. Acad. Eng. Med. Res.* 1, 1–7.

Taggart, R. W., Dressler, M., Kumar, P., Khan, S., and Coppola, J. F. (2016). "Determining emotions via facial expression analysis software," in *Proceedings of Student-Faculty Research Day, CSIS* (New York, NY: Pace University), C2-1–C2-8.

Trevisan, D. A., Bowering, M., and Birmingham, E. (2016). Alexithymia, but not autism spectrum disorder, may be related to the production of emotional facial expressions. *Mol. Autism* 7:46.

Watson, D. A., and Tellegen, A. (1985). Towards a consensual structure of mood. *Psychol. Bull.* 98, 219–235. doi: 10.1037/0033-2909.98.2.219

Weiss, H. M., and Cropanzano, R. (1996). Affective events theory: a theoretical discussion of the structure, causes, and consequences of affective experiences at work. *Res. Organ. Behav.* 18, 1–74.

Wolf, K. (2015). Measuring facial expression of emotion. *Dial. Clin. Neurosci.* 17, 457–462.

Xefteris, S., Doulamis, N., Andronikou, V., Varvarigou, T., and Cambourakis, G. (2016). Behavioral biometrics in assisted living: a methodology for emotion recognition. *Eng. Technol. Appl. Sci. Res.* 6, 1035–1044. Available online at: https://etasr.com/index.php/ETASR/article/view/634/366

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

frontiers in Human Neuroscience

TECHNOLOGY AND CODE
published: 21 April 2020
doi: 10.3389/fnhum.2020.00128

Species Classification for Neuroscience Literature Based on Span of Interest Using Sequence-to-Sequence Learning Model

Hongyin Zhu[1,2], *Yi Zeng*[1,2,3,4]*, *Dongsheng Wang*[5] *and Cunqing Huangfu*[1]

[1] *Research Center for Brain-Inspired Intelligence, Institute of Automation, Chinese Academy of Sciences, Beijing, China,* [2] *School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China,* [3] *Center for Excellence in Brain Science and Intelligence Technology Chinese Academy of Sciences, Shanghai, China,* [4] *National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Science, Beijing, China,* [5] *Department of Computer Science, University of Copenhagen, Copenhagen, Denmark*

OPEN ACCESS

Edited by:
Victor Hugo C. de Albuquerque, University of Fortaleza, Brazil

Reviewed by:
Deepak Gupta, Maharaja Agrasen Institute of Technology, India
Ali Hassan Sodhro, Sukkur Institute of Business Administration, Pakistan
Bo Zheng, Harbin Institute of Technology, China

***Correspondence:**
Yi Zeng
yi.zeng@ia.ac.cn

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

Received: *09 February 2020*
Accepted: *19 March 2020*
Published: *21 April 2020*

Citation:
Zhu H, Zeng Y, Wang D and Huangfu C (2020) Species Classification for Neuroscience Literature Based on Span of Interest Using Sequence-to-Sequence Learning Model. Front. Hum. Neurosci. 14:128. doi: 10.3389/fnhum.2020.00128

Large-scale neuroscience literature call for effective methods to mine the knowledge from species perspective to link the brain and neuroscience communities, neurorobotics, computing devices, and AI research communities. Structured knowledge can motivate researchers to better understand the functionality and structure of the brain and link the related resources and components. However, the abstracts of massive scientific works do not explicitly mention the species. Therefore, in addition to dictionary-based methods, we need to mine species using cognitive computing models that are more like the human reading process, and these methods can take advantage of the rich information in the literature. We also enable the model to automatically distinguish whether the mentioned species is the main research subject. Distinguishing the two situations can generate value at different levels of knowledge management. We propose SpecExplorer project which is used to explore the knowledge associations of different species for brain and neuroscience. This project frees humans from the tedious task of classifying neuroscience literature by species. Species classification task belongs to the multi-label classification which is more complex than the single-label classification due to the correlation between labels. To resolve this problem, we present the sequence-to-sequence classification framework to adaptively assign multiple species to the literature. To model the structure information of documents, we propose the hierarchical attentive decoding (HAD) to extract span of interest (SOI) for predicting each species. We create three datasets from PubMed and PMC corpora. We present two versions of annotation criteria (mention-based annotation and semantic-based annotation) for species research. Experiments demonstrate that our approach achieves improvements in the final results. Finally, we perform species-based analysis of brain diseases, brain cognitive functions, and proteins related to the hippocampus and provide potential research directions for certain species.

Keywords: brain science, neuroscience, cognitive computing, multi-label classification, corpus annotation, PubMed, linked brain data

1. INTRODUCTION

Managing neuroscience literature from species perspective is an innovative and important research task for understanding the functionality and structure of the brain. Species information in scientific works can be used to organize knowledge facts in the Linked Brain Data[1] (LBD) (Zeng et al., 2014b) scheme, and then the system composed of brain and neuroscience communities (Ascoli et al., 2007; Gardner et al., 2008; Imam et al., 2012; Sunkin et al., 2012; Larson and Martone, 2013; Poo et al., 2016), neurorobotics, and other devices can automatically utilize species knowledge on the Internet by accessing the API provided by the LBD platform. For example, brain science knowledge of different species can be used to build brain simulation cloud computing platforms for different animals (Liu et al., 2016), monkey brain-inspired neurorobotics (Zeng et al., 2018), Drosophila brain-inspired Unmanned Aerial Vehicle (UAV) (Zhao et al., 2018), neuroimaging (Zeng et al., 2014a), and help neuroscientists design biological experiments (Poo et al., 2016). Internet of Things for brain science aims to link the brain-related data and devices to the Internet and help research and protect the brain. Our research opens up new opportunities for understanding and exploring the brain of different species to promote brain and neuroscience research. The species classification task is to assign pre-defined species labels to neuroscience literature that does not explicitly mention the species. This technology can be used to classify and organize neuroscience literature based on the species to help researchers and devices easily compare the similarities and differences between different species for linking the brain and neuroscience communities and different devices. The knowledge about certain species can also help find solutions to address some of the major health problems in humans, e.g., the HIV (Micci and Paiardini, 2016), the Jenner vaccine (Riedel, 2005), the Parkinson's disease (Bailey, 2006), etc.

The use of model organisms for human research purposes is commonplace—researchers can study these organisms in ways that are unethical or impractical in humans. Model organisms represent the species that have been extensively studied to understand specific biological phenomena and are usually easy to maintain and breed in a laboratory setting. In this paper, as an illustrative example, we focus on 23 types of representative animal models selected from Neuromorpho.org, i.e., "Agouti, Blowfly, Elegans, Cat, Chicken, Cricket, Dragonfly, *Drosophila melanogaster*, Elephant, Frog, Goldfish, Guinea pig, Human, Monkey, Moth, Mouse, Rabbit, Rat, Salamander, Sheep, Spiny lobster, Turtle, Zebrafish". Many scientific works do not explicitly mention research species, which poses challenges for large-scale automated species extraction and analysis. Although some species can be inferred by manual reading and analysis of other information in the literature, such as target gene terms, organs, and functions, it is already difficult for humans to read a hundred articles. Analyzing millions of literature in this way is almost impossible. When classifying these documents, the human brain uses not only the brain's dictionary matching mechanism but also other mechanisms (such as attention and memory). The secondary challenge is how to guess various species at once. The research of other species is crucial for the study of brain and neuroscience. Faced with large-scale literature, it is inefficient to manually summarize species or to infer species using complex processes.

Species information is one of the most basic information that researchers are concerned about. (1) Researchers based on model organisms first focus on what species the research is based on. Because the species studied in the paper determine whether this paper has reference value or impact on their research. When research problems shift from frontier species to later species, a lot of species matching work is needed. It would be great if the species could be identified automatically. For example, specific genes related to working memory have been studied in *Drosophila melanogaster*, and they have also been found in mice, but no experiments have been performed. If the researcher doing the mouse experiment wants to search all the genes that have been studied in other species, or if he wants to search whether the specific genes present in mice have been studied in other species, then he first needs to know which species were studied in each article. Species are important information in biological research because each species has different characteristics, the research area suitable for each species is different, and the infrastructure investment (e.g., smart animal house, humidity and temperature control devices, laboratory instrument, etc.) of each species is also different. For example, zebrafish are suitable for exploring developmental problems, and fruit flies are more likely to perform genetically modified experiments. It is difficult to use mice to study developmental problems. It is important and instructive to make full use of species information for knowledge integration. (2) For researchers who do not consider too much species information, they also need to be aware of the importance of species in their research. If researchers want to write a review, such as a survey of mice or fruit flies, the need to use such a toolkit to eliminate many unnecessary papers. (3) If researchers want to build an automated literature analysis system in a certain field, the lack of species information will lead to confusion of knowledge on the Internet. In subsequent applications, users cannot get the results they are searching for. Machines simply cannot distinguish which species the knowledge belongs to, so this system cannot be easily accomplished.

Brain science knowledge urgently needs to be managed from a species perspective. Otherwise, this knowledge will be mixed, which will seriously affect subsequent applications and elements, including biologists/researchers who perform literature analysis and the automated literature analysis systems on the Internet. We need to use the knowledge of other species to solve the problems of humans. Categorizing several documents manually does not yield much valuable information. Categorizing large-scale literature by species will help harness the knowledge of other species to solve the problems of humans. This paper proposes a framework that can effectively process large-scale documents, improves the efficiency of literature analysis, and organizes the brain science knowledge based on species of interest. This framework uses not only species mentions and genetic terms but also cognitive computing models to process the contextual expressions and span of interest in the text. Our work has greatly

[1] http://www.linked-brain-data.org

improved the efficiency of species analysis and data transmission on the Internet.

This task can be formulated as two different task schemes, the text classification scheme (discriminative model) and the text summarization scheme (generative model). The text classification scheme classifies a document into different species, while the text summarization scheme summarizes the document from a species view and naturally considers the label correlation. The text classification scheme is easier because a document can be encoded as a fixed-length vector to retain the main information. The challenge is how to emphasize effective information about species in a long document. Note that this is a multi-label classification (MLC) task since a scientific work may be related to two or more species. The text summarization scheme is more like the human reading process because when humans read the paper, we gradually discover each species by mapping to different parts of the paper. Although the labels are obtained in a certain order, this order is not considered in evaluation—and this is not needed, as it is being used as a MLC problem. Inspired by the human reading process, the text summarization model gradually generates each species by attending to the span of interest (SOI) and considers the correlation between the tags. SOI in text is equivalent to region of interest (ROI) (Girshick et al., 2014; Girshick, 2015; Ren et al., 2015; He et al., 2017) in a picture. ROI is widely used in object detection of computer vision (CV) and it can be any particular portion of the image that seems important for the task. Here, we use SOI to represent the important text spans for species prediction.

The PubMed[2] provides the citations of references and abstracts of biomedical literature from MEDLINE, life science journals, and online books. The PubMed Central (PMC)[3] archives publicly accessible full-text articles of biomedical and life sciences journal literature. The research project of this paper is mainly about knowledge linking and extraction in the field of brain and neuroscience. Linked Brain Data (Zeng et al., 2014b, 2016; Zhu et al., 2016b,c) is an effort for extracting, integrating, linking and analyzing brain and neuroscience data and knowledge from multiple scale and multiple data sources. This platform focuses on the associations among brain regions, brain diseases, cognitive functions, neurons, proteins, and neurotransmitters. There are more than 2,339,898 relational triples in the LBD platform, such as (Hippocampus, relatedTo, Alzheimer's disease), (Hippocampus, relatedTo, Associative memory). These relations are machine-readable structured knowledge. This paper can organize massive structured brain science knowledge according to different species, thereby forming the structured species knowledge, which can be considered as 4-ary, e.g., (Hippocampus, relatedTo, Alzheimer's disease, Human) or (Hippocampus, relatedTo, Alzheimer's disease, Monkey). The proposed approach can facilitate the cross-species brain science research. The LBD platform provides services to connect the brain and neuroscience communities and devices.

A commonly used multi-label approach is the binary method (Fan and Lin, 2007) which builds a decision function for each class. Despite the success of the MLC scheme, it is often necessary to find a threshold to convert the probability value into a true/false flag for each class so that we can select a subset of the species as the final result. The thresholds for different species are usually different, and the final result is affected by the hard threshold. Finding globally optimal thresholds (Fan and Lin, 2007) for all classes is complicated. Inspired by Yang et al. (2018), we propose the sequence-to-sequence classification (SeqC) framework. Different from the MLC scheme, our SeqC framework does not need to search the thresholds because each step only outputs the most probable label by emphasizing SOIs. When there are no more species, this model will output the stop tag (Bahdanau et al., 2015). Abstractive summarization models usually have a ground truth sequence to learn how to paraphrase the main content of the passage and may use the teacher forcing (Williams and Zipser, 1989) and the scheduled sampling (Bengio et al., 2015) to improve the model performance. In contrast, this task only has class labels without the sequence order, so we convert species labels into virtual species sequences in a fixed order. During the model evaluation, we do not consider the label order.

MLC is more complex than single-label classification in that the labels tend to be correlated and different parts of a document have different contributions when predicting labels. Our decoder considers the correlations between species by processing species dependencies through LSTM units. A document can be very long, which poses a challenge for the one-level encoding model. Besides, not all sentences help to predict the species and not all words contribute equally to a sentence. To solve these two problems, we integrate the hierarchical document encoding and hierarchical attentive decoding (HAD) into the sequence-to-sequence model. We consider the word- and sentence/section-levels. Besides, simple MLC models only generate a vector representation that calculates an attention distribution over the document. Different species are usually associated with different parts of the document, so simple MLC models cannot adaptively attend to different parts of the document for different species, which potentially limits the performance. In contrast, our sequence-to-sequence classification model allows each species prediction to attend to different parts of the document.

To train and evaluate models, we label the PubMed and PMC corpora[4]. We present two versions of annotation criteria (mention-based annotation and semantic-based annotation). This paper is organized below. Section 3 describes the core modules of this framework. Section 4 describes the labeled datasets and experimental analysis. The major contributions of this paper can be summarized below.

I. This paper formulates a new task, species classification in neuroscience literature. We propose the SeqC framework to classify neuroscience literature based on SOIs. This study improves the transfer efficiency of brain science knowledge

[2]https://www.ncbi.nlm.nih.gov/pubmed/
[3]https://www.ncbi.nlm.nih.gov/pmc/
[4]https://github.com/sssgrowth/SPECIESEXPLORER

on the Internet and opens up opportunities for brain science text mining from the species perspective.

II. Our approach integrates the hierarchical document modeling and hierarchical attentive decoding to model the document structure and extract informative SOIs related to species. This framework supports both dictionary-based method and various deep learning models.

III. We create three datasets which label 23 types of representative species in the PubMed and the PMC corpora. We propose two versions of annotation standards to facilitate the use of knowledge extraction in brain science text mining. This process is semi-automated and easily extendable to greater sets of species.

2. RELATED WORK

Some works use the knowledge of different animals to resolve biological and biomedical questions. The species information can be used to manage the facts in a knowledge base to support the research of brain and neuroscience, such as the Brain Knowledge Engine[5] (Zhu et al., 2016a). They organize the knowledge with species meta-data and explore the multi-scale nervous systems, cognitive functions and diseases of different species for linking brain and neuroscience communities, neurorobotics, brain simulation cloud computing platform, and other devices on the Internet by accessing the API. Norouzzadeh et al. (2018) propose a method to identify the location and behavior of animals from pictures to study and conserve ecosystems. McNaughton et al. (1983) study the contributions of position, direction, and velocity to single unit activity in the hippocampus of rats. Leach et al. (1996) found that blockade of the inhibitory effects of CTLA-4 can allow for, and potentiate, effective immune responses against tumor cells on mice. The above two contributions won the Nobel Prizes in Medicine because they have profound implications on human biomedical research. The animal information is also helpful for the study of the welfare of the animals, and the concept of animal rights (Andersen and Winter, 2017).

The technologies for the Internet of Things (Gochhayat et al., 2019; Kumar et al., 2019; Bebortta et al., 2020; Qian et al., 2020) are also widely used in different domains for understanding the functionality and structure of the brain and address some problems in human daily life. De Albuquerque et al. (2017) investigate the applications of brain computer interface systems. Some IoT frameworks are proposed to analyze the brain signals, such as brain CT images (Jaiswal et al., 2019; Sarmento et al., 2020; Vasconcelos et al., 2020), MRI (Mallick et al., 2019; Arunkumar et al., 2020), etc. Many applications benefit human daily life. Innovative algorithms for improving video streaming are proposed in the Internet of Multimedia Things (IoMT) and Internet of Health Things (IoHT) to optimize the Telemedicine and medical quality of service (m-QoS) (Sodhro et al., 2018). Sodhro et al. (2019a) propose the QGSRA algorithm to alleviate fluctuation in the wireless channel to support multimedia transmission. Using artificial intelligence algorithms to solve accurate resource management and energy efficiency issues (Sodhro et al., 2017, 2019b) is an important aspect of implementing the Internet of Things.

The NCBI Taxonomy[6] (Federhen, 2011) is a curated classification and nomenclature for all of the organisms in the public sequence databases. It accounts for about 10% of the described species of life on the planet. It includes more than 234,991 species with formal names and another 405,546 species with informal names. Currently, the experiments of this paper focus on the 23 model organisms because there are systematic research methods for these species. Bada et al. (2012) create the Colorado Richly Annotated Full-Text (CRAFT) Corpus which contains 97 articles and annotates the concepts from 9 well-known biomedical ontologies and terminologies. Funk et al. (2014) evaluate dictionary-based concept recognizers on eight biomedical ontologies in the CRAFT dataset. Biomedical natural language processing (BioNLP) (Ananiadou and McNaught, 2006; Cohen and Demner-Fushman, 2014; Wei et al., 2015) aims to enable computers to efficiently read the vast amount of the literature and extract key knowledge about specific topics. There are some BioNLP tasks and corpora in the context of the BioCreative and BioNLP shared tasks. BioNLP (open) shared tasks (Dubitzky et al., 2013) contains a series of computational tasks of biomedical text mining (TM), evaluations, and workshops. Critical Assessment of Information Extraction in Biology (BioCreative) (Hirschman et al., 2005; Hemati and Mehler, 2019) includes assessments of biological domain information extraction and text mining development across the community.

BioNLP has achieved substantial progress on many tasks (Ananiadou and McNaught, 2006; Hunter and Cohen, 2006; Jensen et al., 2006), such as named entity recognition, information extraction, information retrieval, corpora annotation, evaluation, etc. These researches open up opportunities to integrate biomedical text mining with knowledge engineering and data mining. Many NLP techniques can be used to extract linguistic features from text in different languages for model learning, such as part-of-speech tagging, word segmentation, linguistic parsing (Manning et al., 2014; Zheng et al., 2016; Che et al., 2018; Li et al., 2019; Wang et al., 2020), etc. There are some researches on text mining in the genomics domain (Zweigenbaum et al., 2007), e.g., identifying gene/protein names and their relations. Hersh (2008) introduce the methods and challenges in many aspects of health and biomedical information retrieval systems. Bodenreider (2008) describe the role of biomedical ontologies in knowledge management, data integration, and decision support. There are some ontologies, such as SNOMED CT, the Logical Observation Identifiers, Names, and Codes (LOINC), the Foundational Model of Anatomy, the Gene Ontology, RxNorm, the National Cancer Institute Thesaurus, the International Classification of Diseases, the Medical Subject Headings (MeSH), and the Unified Medical Language System (UMLS). Smith et al. (2007) introduce the shared principles governing ontology development in the Open Biomedical Ontologies (OBO). Curtis et al. (2005), Khatri and Drăghici (2005), and Huang et al. (2008) use microarray

[5]http://www.brain-knowledge-engine.org

[6]https://www.ncbi.nlm.nih.gov/taxonomy

technology and Gene Ontology (GO) terms to analyze the gene expression to characterize biological processes and identify the mechanisms that underlie diseases.

A commonly used multi-label approach is the binary method, which constructs a decision function for each class. Fan and Lin (2007) present a method to adjust the decision thresholds for each class. Zhang and Zhou (2007) propose the BP-MLL with a fully-connected neural network and a pairwise ranking loss function. Kim (2014) proposes the one layer CNN architecture with multiple filter width to encode both task-specific and static vectors. Nam et al. (2014) propose a neural network using cross-entropy loss instead of the ranking loss. Kurata et al. (2016) utilize word embeddings based on CNN to capture label correlations. Yang et al. (2016) propose a hierarchical attention network (HAN) to encode the sentence representation and document representation. They experimented with IMDB reviews, Amazon reviews, etc. for sentiment estimation and topic classification (Di Buccio et al., 2018; Tiwari and Melucci, 2018a,b, 2019a,b). Our model also considers the hierarchical attention, but the difference is that our model uses a decoder to resolve the multi-label classification problem and to calculate the hierarchical attention. Our proposed method uses the HAD mechanism in the decoder for each species prediction, while HAN calculates the attention in the encoding process. Besides, our model considers the discourse sections structure in scientific works during the decoding process. Liu et al. (2017) present a variant of CNN based approach to extreme multi-label text classification. Chen et al. (2017) propose a method to ensemble the CNN networks to capture diverse information on different nets. See et al. (2017) present the pointer generator network for text summarization. Yang et al. (2018) propose a sequence generation model for MLC. Cohan et al. (2018) propose a discourse-aware attention model for text summarization. They consider each section as a sequence and attending to the sequences of words. Inspired by the above studies, we integrate hierarchical document modeling, sequence-to-sequence model, and HAD into our species classification model.

3. METHODS

First, we give an overview of the model. Second, we describe data acquisition, processing, and corpus annotation of the PubMed and PMC literature. Then, we explain in detail the SeqC framework of encoder and decoder which includes the sequence-to-sequence scheme and the hierarchical attentive decoding mechanism. Finally, we introduce the training method.

3.1. Overview

First, we define some notations and describe the species classification task. Given the predefined m species $L = \{c_1, c_2, ..., c_m\}$ and a scientific work (neuroscience literature), our model assigns a subset of species to this document. More formally, each document has a list of predefined species candidates $\{y_1, y_2, ..., y_m\}$, where the label of the i-th species (c_i) is $y_i \in \{0, 1\}$ with 1 denotes a positive class and 0 otherwise. Our goal is to learn a model that can select the possible species subset involved in this scientific work. From the perspective of sequence-to-sequence model, this task can be modeled as finding an optimal species combination y^* that maximizes the conditional probability $p(y|x)$, which is calculated as follows.

$$p(y|x, \theta) = \prod_{i=1}^{m} p(y_i|y_1, y_2, ..., y_{i-1}, x, \theta) \quad (1)$$

where θ is the model parameter. The loss of the whole dataset can be calculated as Equation (2). We sort the label sequence of each sample according to the label frequency in the training set, with the higher frequency labels ranked front. For multi-label classification problems, the order of the labels is not needed for the result evaluation. We tested several methods to sort the labels and found that the results were almost the same.

$$L(\theta) = \sum_{j} p(y^j|x^j, \theta) \quad (2)$$

where j is the j-th document.

$$y^* = \arg \max_{y \in Y(z)} \log p(y|x, \theta) \quad (3)$$

where $Y(z)$ denotes 2^m possible combinations.

An overview of our proposed model is shown in **Figure 1**. Our main effort lies in designing a model that predicts each species by emphasizing SOI from the document. First, we convert the ground truth label into a species combination sequence. This allows the model to predict each species sequentially. Besides, the beginning symbol (BOS) and end symbol (EOS) are added to the head and tail of the species labels, respectively. Second, we use the two-level encoder to generate the contextual representation of the sentence/section and the document respectively. Finally, the decoder predicts each species by using the HAD mechanism.

This model can be seen as a simplified version of the neural abstractive text summarization model. Text summarization has a larger vocabulary for summarizing the main content, while the size of our vocabulary is 23. Text summarization allows the same words appear repeatedly in the output, while in our model each class label only appears once, so it reduces the repetition problem (See et al., 2017) in text summarization. Text summarization has the problem of out-of-vocabulary (OOV) words and uses the copy mechanism (See et al., 2017) to solve it, while our model does not have this problem since all the labels are fixed. In summary, this approach is promising in this task since this task is well-defined under the sequence-to-sequence classification scheme.

3.2. Data Processing

3.2.1. Data Acquisition and Preprocessing

To obtain the neuroscience literature, we download all biomedical literature from 1987 to 2019 on the PubMed[7] and PMC[8]. Then, we retrieve the biomedical literature related to

[7]https://ftp.ncbi.nlm.nih.gov/pubmed/
[8]https://ftp.ncbi.nlm.nih.gov/pub/pmc/

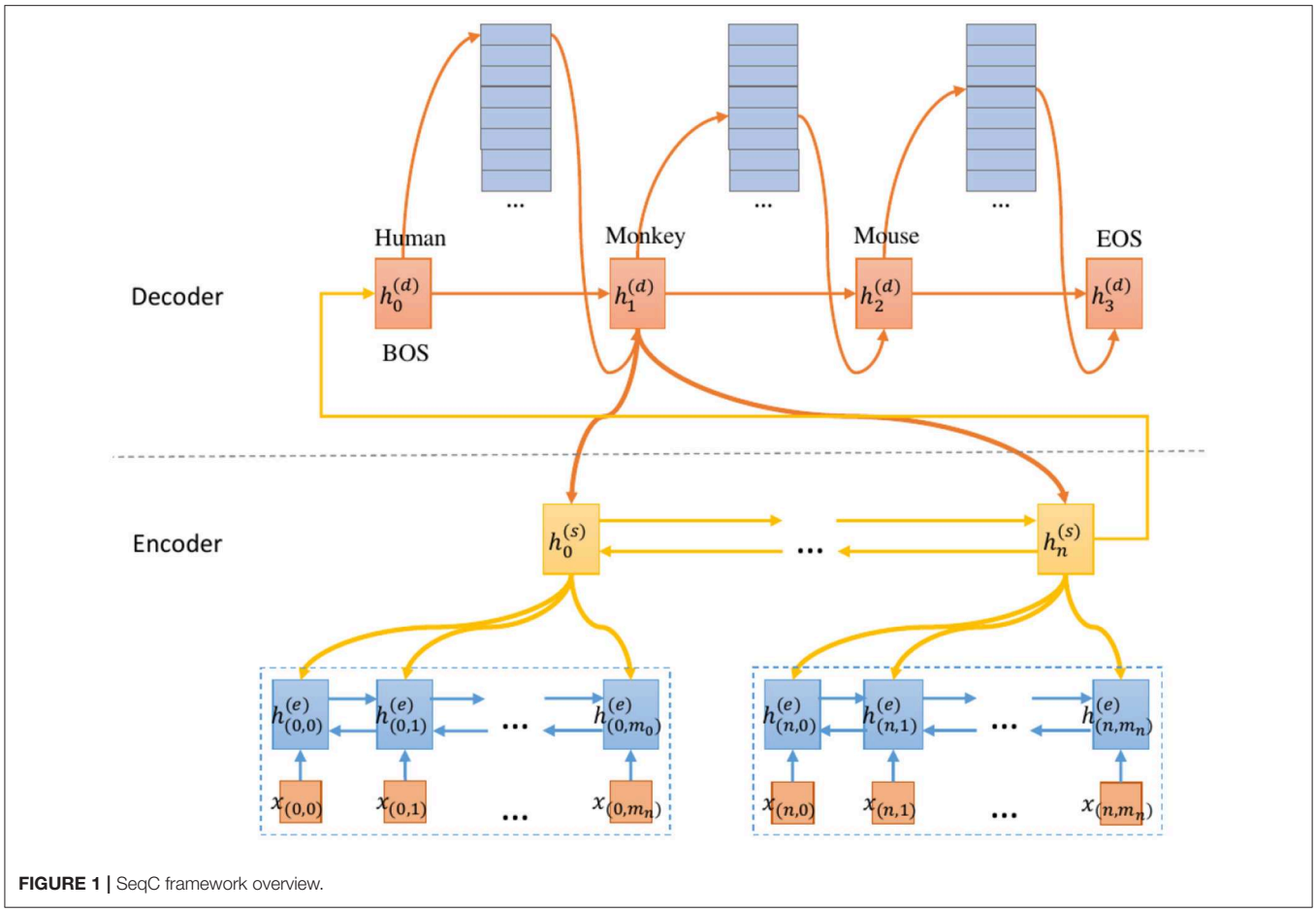

FIGURE 1 | SeqC framework overview.

neuroscience. We tokenize the documents and match the case-insensitive prefix (i.e., *brain,neuron,neural,neuro,cerebral*) at the word level.

In order to reduce the impact of the references and additional sections, we analyze the XML tag name and use the regular expression to extract the PMID/PMC, article title, abstract, keywords, article body, and date. We deleted tables to only preserve the textual content. We also convert XML escape characters into human-readable characters, for example, converting < to <, > to >, & to &, " to", etc. Then we select the literature by matching the keywords in the title, abstract, and body of the article.

3.2.2. Full-Category Sampling

We sample two sets of documents from PubMed and PMC corpora respectively. The set of articles in the PubMed corpus overlaps with the articles in the PMC corpus, given that the PMC articles would have a corresponding abstract in PubMed. To make the two datasets independent of each other, we removed the overlapping abstracts. The PubMed dataset contains 5,040/778/775 documents as the division of training/development/test (train/dev/test) sets. The PMC corpus contains 1,427/204/195 documents. In order to make the dataset cover all categories and better reflect the distribution of categories, we propose the full-category sampling (FCS) algorithm, as shown in Algorithm 1.

During the sampling process, we shuffle the documents and randomly select 50,000 documents as candidate documents. If the class support degree of species x (e.g., Mouse) reaches 400, this method no longer samples this species. The x denotes any pre-defined species. This class support degree denotes the maximum number of documents in each class. This method ensures that the dataset can cover all categories. The key insight of this algorithm is that it can prevent the oversampling of sparse classes.

We explain this algorithm. As shown in line 1, this method shuffles the corpus and randomly samples the candidate set. This operation prevents the oversampling of sparse classes. Otherwise, for sparse classes, this method will skip too many unrelated documents until enough samples of this class are obtained. Then, we initialize the *specDict* and *samples* to hold the sample results. Note that each sample is annotated with the mention-based annotation described in subsection 3.2.3. In lines 4–14, if the tag of the *i*-th document contains species x and the number of documents related to species x does not reach the class support degree *s*, the *i*-th document will be added to the dataset. Finally, *samples* contains the selected documents.

Algorithm 1: The full-category sampling algorithm

Require: The corpus D with species labels for each document, class support degree s, candidates number n
Ensure: The sampled dataset *samples*

```
Shuffle the corpus D and sample n candidates D'
specDict = {}, samples = []
for i ← 0, D'.length − 1 do
    doc, tags = D'[i]
    added = False
    for j ← 0, tags.length − 1 do
        if !specDict.contains(tags[j]) then specDict[tags[j]]=0
        end if
        if specDict[tags[j]] < s then
            specDict[tags[j]]++
            if added != True then Add D'[i] to samples
            end if
            added=True
        end if
    end for
end for
```

3.2.3. Corpus Annotation

From the perspective of literary expression, the expression of related species is mainly divided into two types. First, some species are mentioned in the literature, such as monkeys, but monkeys themselves are not the main experimental subjects. Monkeys are associated with this study. This information can help find more comprehensive and instructive relevant knowledge. Second, this species is the main experimental subjects of the literature. This information can produce accurate semantic search results. Both cases have high research value. We create two versions of the dataset which are the mention-based annotation and the semantic-based annotation.

3.2.3.1. Mention-based annotation

The first version (mention-based annotation) follows the criteria of species mention, which considers all the mentioned species as labels. More formally, let $c_i \in C$ denote a predefined species, where C is the pre-specified species set. $s_j \in S$ is a sample (i.e., an abstract or an article). If s_j mentions c_i (including one of its synonyms, variants, subspecies and its common alias from NCBI Taxonomy vocabulary), we assign c_i to s_j. We consider the singular and plural forms of the species. We use the above dictionary-based method to label the entire dataset. Labeling documents that explicitly mention species is straightforward and efficient. The advantage is that it can find more relevant and comprehensive species to a study. After that, we can use these species labels as keys to efficiently retrieve the literature related to a specific species. This process avoids repeated computation and saves resources. The species tags of each article link massive documents. Users can utilize species tags to get more articles. This method is more complete and efficient than using words to retrieve plain text.

We also let three human annotators check the comprehensiveness and correctness of the species labeled for each sample. For example, some documents use other words related to humans, e.g., "humankind, humanity, humane, man, woman, men, women, male, female, patients." Overview articles also follow this annotation standard consistently, so they are considered relevant to the species mentioned. A conclusive dataset is generated using the combination of these annotations by an independent person.

The dictionary-based method may not perform well in the following situations. Sometimes, it is necessary to use context to determine whether "cricket" is a species or a game and whether "mouse" is an animal or a computer device. There are 18 PMC articles and 2 PubMed abstracts use "cricket" as the game. For example, "Hamstring injuries are not confined strictly to Australian Rules football but are also seen in soccer, athletics, hurling, **cricket** and touch football (Hoskins and Pollard, 2005)." There are 6 PMC articles and 1 PubMed abstract use "mouse" as the computer device. For example, "Total in-home computer use per day was calculated using **mouse** movement detection and averaged over a 1-month period surrounding the MRI (Silbert et al., 2016)." The weakness is that this standard may introduce some noisy species labels when they are not the main research subjects of the literature. This problem can be resolved by the following semantic-based annotation.

3.2.3.2. Semantic-based annotation

The second version (semantic-based annotation) follows the criteria of expert knowledge. We let domain experts in the field of biology manually label the above PMC dataset based on the main research subjects of the article body. However, this process is costly and time-consuming, because annotators need to read the article and discuss the annotation standard. We add "cell," "not applicable," and "others" classes in that most cell-centric experiments share common methodologies. It is valuable to consider the "cell" as a class. For example, there are a lot of drug tests on cell or expression system related researches. Besides, a few papers did not study these species. We also need to use appropriate levels of species as the label to generate more valuable information. For the moth, considering a specific moth cannot generate much valuable information. The advantage of this standard is that articles retrieved using the primary research subject are more likely to contain satisfactory knowledge. However, the weakness is that the recall may not be high enough. For example, humans are not actually studied in some articles, but the research as a whole is done for the purpose of gaining insight into a disease that affects humans. There are 968 such documents without human labels. The mention-based annotation can make up for this problem. The mention-based annotation generally mine more species from these documents. Detailed standard is described in section 1 in the **Supplementary Material**[9].

3.2.3.3. Inferring species from the literature

To evaluate whether our model can infer species from the literature that does not mention species, we hid the

[9]https://github.com/sssgrowth/SPECIESEXPLORER/blob/master/icon/appendix.pdf

species mentions and substituted them with the same symbol "*SPECIES*" to simulate the document that does not mention species. For example, masking "monkey" and "mouse" in a document (Cho et al., 2019), the sentence

> We have established monkey NPC cell lines from induced pluripotent stem cells (iPSCs) that can differentiate into GABAergic neurons *in vitro* as well as in mouse brains without tumor formation.

becomes

> We have established *SPECIES* NPC cell lines from induced pluripotent stem cells (iPSCs) that can differentiate into GABAergic neurons *in vitro* as well as in *SPECIES* brains without tumor formation.

Masked language models predict each masked token in the sentence, which is the token-level prediction. Different from the masked language model, we do not predict the masked token in the document, instead we predict each species only once and the prediction happens in the whole document, which is the document-level prediction. Masking species enables the model to learn how to use other information in the text to execute inference. Otherwise, the attention focuses on species words, not generating much valuable information. Besides, the performance of all models on the PubMed dataset is almost the same as using a dictionary-based method. In practice, this model does not need the above mask operation since we can input the original scientific work (with or without mentioning the species). To quantitatively analyze the inference performance, this way of data creation can reduce the risk of missing species. We also test our model when restoring the species mention. We keep the original files for human access. This would be critical for correct resolution.

3.3. Encoder

Our encoder extends the RNN encoder to the hierarchical RNN that captures the document structure. We first encode each sentence/section and then encode the document. The word-section level encoding is only used to model the article body. The abstract does not have section, but we unify these two modeling into one framework. Therefore, $h_i^{(s)}$ denotes sentence and section interchangeably. Formally, we encode the document as a vector based on the following equation:

$$h^{(doc)} = RNN_{doc}(h_1^{(s)}, h_2^{(s)}, ..., h_n^{(s)}) \tag{4}$$

$RNN(\cdot)$ represents a recurrent neural network whose final state is used to represent the input sequence. n is the number of sequences in the document. The superscript $^{(s)}$ and $^{(doc)}$ denote the sentence/section and the document representation respectively. $h_i^{(s)}$ is the representation of the i-th sequence, which is computed as follows.

$$h_i^{(s)} = RNN_s(x_{(i,1)}, x_{(i,2)}, ..., x_{(i,m)}) \tag{5}$$

where $x_{(i,j)}$ is a word embedding of token $w_{(i,j)}$ and m is the sequence length. The parameters of $RNN_s(\cdot)$ are shared by all the sentences/sections. We use the single layer bidirectional LSTM for both $RNN_{doc}(\cdot)$ and $RNN_s(\cdot)$ to encode hidden states.

3.4. Decoder

3.4.1. Sequence-to-Sequence Scheme

See et al. (2017) present the pointer-generator network for text summarization. Different from them, our decoder aims to model the correlation between species. At each step t, the decoder (a single-layer unidirectional LSTM) receives the species embedding of the previous step and the information of the input document. During training, the previous species comes from the ground truth label; at test time, the previous species is emitted by the decoder. The hidden state $h_t^{(d)}$ at time step t is computed as follows.

$$h_t^{(d)} = RNN_{dec}([spec(y_{t-1}); c_{t-1}], h_{t-1}^{(d)}) \tag{6}$$

where [;] denotes the concatenation operation. The superscript $^{(d)}$ denotes the decoder. $RNN_{dec}(\cdot)$ is a uni-directional LSTM-RNN decoder. $spec(y_{t-1})$ denotes the species embedding with the highest probability under the prediction distribution y_{t-1}. y_{t-1} is the prediction of the previous step. c_{t-1} is the context vector generated from the input document using the hierarchical attention mechanism. $spec(y_0)$ is initialized to a trainable vector. c_0 and $h_0^{(d)}$ are initialized to a zero vector and the document vector $h^{(doc)}$ respectively.

3.4.2. Hierarchical Attentive Decoding Mechanism

When the model predicts certain species, not all sentences/sections and words contribute equally. The attention mechanism can generate a context vector by attending to the SOIs of the document and aggregating their contextual representations. Modeling an article directly into a sequence of words cannot fully preserve the information and structure of the document. Discourse structure (Tang et al., 2015) information has proven effective in modeling document. Scientific works are usually composed of standard discourse sections structure describing the problem, methodology, experiments, conclusions, etc. Cohan et al. (2018) present a discourse-aware attention mechanism that generates better representation by incorporating discourse sections structure knowledge in the model architecture. We propose the HAD mechanism to consider discourse sections information for species prediction so that the model can extract important information from the literature more accurately based on the discourse sections, thus obtaining a better vector representation. Most literature only provides abstracts, so we use the HAD mechanism for the word and the sentence/section. When we process the full-text, our model uses the discourse sections structure, like (Cohan et al., 2018).

Specifically, the context vector related to the species information is computed as follows.

$$c_t = \sum_i^n \sum_j^m \alpha_{t(i,j)} h_{(i,j)}^{(e)} \tag{7}$$

where $h^{(e)}_{(i,j)}$ is the hidden state of the encoder for the j-th word in the i-th section. The superscript $^{(e)}$ denotes the encoder. $\alpha_{t(i,j)}$ denotes the attention weight of the j-th word in the i-th section at the t-th step. The scalar weight $\alpha_{t(i,j)}$ is computed as follows.

$$\alpha_{t(i,j)} = \underset{(i,j)}{\text{softmax}}(\beta_{t(i)}\text{score}(h^{(e)}_{(i,j)}, h^{(d)}_{t-1})) \tag{8}$$

where the score(·) function is the additive attention function, as shown in formula (10). $\beta_{t(i)}$ is the weight of the i-th section at the t-th step. We parse the start and end positions of each section from the original literature files using the DOM parser so that we can find discourse sections.

$$\beta_{t(j)} = \underset{i}{\text{softmax}}(\text{score}(h^{(s)}_{i}, h^{(d)}_{t-1})) \tag{9}$$

The correlation score is calculated by the additive attention (Bahdanau et al., 2015). $h^{(s)}_i$ denotes the hidden state of the i-th section.

$$\text{score}(h^{(e)}_{(i,j)}, h^{(d)}_{t-1}) = v^T tanh(W_1 h^{(e)}_{(i,j)} + W_2 h^{(d)}_{t-1} + b^{(d)}) \tag{10}$$

where $v \in \mathbb{R}^\tau$ is a weight vector. $W_1, W_2 \in \mathbb{R}^{\tau\times\tau}$ are weight matrices. $b^{(d)} \in \mathbb{R}^\tau$ is a bias vector.

3.5. Training Method

At the t-th decoding step, the vector $h^{(d)}_t$ generated by the decoder is used to predict the probability distribution of each class by the softmax function, as shown in Equation (11).

$$\hat{y} = \text{softmax}(Wh^{(d)}_t + b + I_t) \tag{11}$$

where the W and b are the weight matrix and bias vector. $I_t \in \mathbb{R}^m$ is the mask vector that prevents the decoder from predicting repeated species.

$$(I_t)_i = \begin{cases} -\infty, & \text{if species } y_i \text{ has been predicted at previous time steps} \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

At the training time, the objective function is the cross-entropy loss as follows.

$$\min_{\Theta} L = -\sum_{i}^{|\mathbb{D}|} \frac{1}{l^{(i)}} \sum_{t}^{l^{(i)}} y^{(i)}_t \cdot \log(\hat{y}^{(i)}_t) \tag{13}$$

where i is the document index and t is the decoder time step. Θ is the model parameter. $|\mathbb{D}|$ is the size of the training set. $l^{(i)}$ is the decoder sequence length of i-th document. $\hat{y}^{(i)}_t$ is the predicted probability of ground truth class $y^{(i)}_t$ at the t-th time step. At test time, we use the beam search algorithm (Wiseman and Rush, 2016) to find the top-ranked prediction sequence.

4. RESULTS

In this section, we conduct experiments on three datasets. We first introduce the datasets, evaluation metrics, implementation details. Then, we compare our method with baselines. Finally, we analyze the model components and experimental results.

4.1. Experimental Settings

4.1.1. Dataset

4.1.1.1. PubMed

Corpus contains 2.55M abstracts, including 22.9M sentences, related to neuroscience science. 1.21M (47.5%) documents mention at least one pre-defined species using the mention-based annotation. The labels of these documents may not be complete, as the abstract may not mention all species. These documents can be used for further research in knowledge linking and extraction projects. We sample 5,040/778/775 documents as the experimental train/dev/test datasets. **Figure 2A** visualizes the distribution of sentence number of the abstract. The x and y axes are the sentence number in a scientific work and the count of scientific works that have the corresponding number of sentences respectively. Each document averagely contains 8.9 sentences. **Figure 2B** visualizes the sentence length distribution. **Figure 3A** visualizes the species distribution. "Human", "Mouse," and "Rat" are more frequent labels.

4.1.1.2. PMC mention

Corpus consists of 0.43M articles, including 54.3M sentences, related to neuroscience science. 0.36M (83.5%) documents mention at least one pre-defined species. Annotating the entire corpus is costly and time-consuming, so we sample 1,427/204/195 documents as the train/dev/test datasets for our experiments. **Figure 2C** visualizes the distribution of sentence number of the paper. The sentence distribution varies over a wide range (14–3,087). Long documents occupy a small portion, so we merge the documents with more than 600 sentences. The criteria of this corpus is the species mention. Each document averagely contains 205.6 sentences. **Figure 2D** visualizes the sentence length distribution. **Figure 3B** visualizes the species distribution. "Human," "Mouse," "Rabbit," and "Rat" are more frequent labels.

4.1.1.3. PMC semantics

Dataset uses the same documents of the PMC Mention dataset. We let domain experts annotate these documents. The criteria of this version are based on expert knowledge. **Figure 3B** visualizes the species distribution. "Human," "Mouse," "Not applicable," and "Cell" are more frequent labels.

4.1.2. Evaluation

In single-label classification (1-of-n), the prediction can be either correct or wrong. Compared with the single-label classification, MLC is unique since the prediction can be partially correct (Venkatesan and Er, 2014). MLC requires different evaluation metrics to evaluate the partially correct. Following (Zhang and Zhou, 2007; Chen et al., 2017; Yang et al., 2018), we adopt the Hamming loss, micro-F1 score. Besides, we also measure the macro-F1 score and F1 per document. F1 per document would

also be informative to measure document-level performance. This metric is calculated by averaging the precision, recall, and F1 of each document.

$$\text{Hamming} = \frac{1}{|N| \cdot |L|} \sum_{i=1}^{|N|} \sum_{j=1}^{|L|} \text{xor}(y_{(i,j)}, t_{(i,j)}) \tag{14}$$

4.1.2.1. A. Hamming loss

Calculates the fraction of wrong labels. The lower the hamming loss, the better the performance is, as shown in formula (14). For an ideal classifier, the Hamming loss is 0.

$$\text{F1} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{15}$$

4.1.2.2. B. Micro-F1

Is the harmonic mean of micro-precision and micro-recall as formula (15). This metric calculates metrics globally by counting the total true positives, false negatives and false positives. This metric aggregates the contributions of all classes.

4.1.2.3. C. Macro-F1

Computes the metric independently for each class and then take the average. This measurement treats all classes equally. We can evaluate the overall model performance for all classes.

4.1.3. Implementation Details

Table 1 reports the main hyperparameters. We train the 200-D GloVe embedding on the whole PubMed and PMC corpora (3M documents). We did not update the pre-trained word embeddings during model training. For the character embeddings, we initialize each character as a 25-D vector. If using character Bi-LSTM, we set 50-D hidden state. If using character CNN, the convolution kernel width is 3, and we use max-pooling to generate 100-D vector representation. The Bi-LSTM dimension of encoder and decoder is 200-D. We use the Adam algorithm (Kingma and Ba, 2014) to train the model. The initial learning rate is 0.001. The size of species embedding is 200-D. We limit the sentence length to 128 and section length to 512 tokens. We conducted experiments on an Intel(R) Xeon(R) CPU E7-4830 v3 @ 2.10 GHz (Mem: 976G) and the GPU Tesla K40c (12G) and TITAN RTX (24G).

4.2. Baseline Models

We compare our method with several baseline models. The Dictionary-based method uses string matching. To extract more species, the glossary of species includes species names, synonyms, variants, subspecies, and its common alias.

The LSTM (Zhang et al., 2015) and CNN (Kim, 2014) models consider the document as a sequence of words and generate a

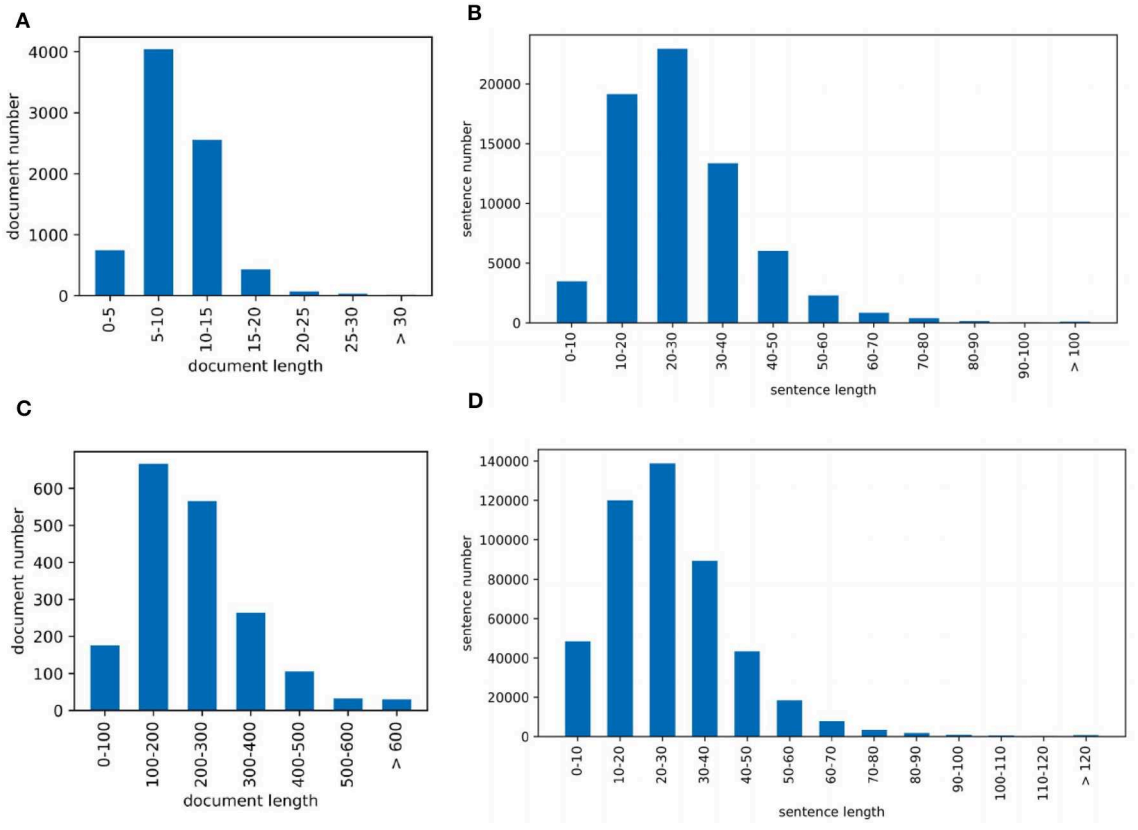

FIGURE 2 | Dataset visualization where **(A)** is the PubMed sentence distribution of each document and **(B)** is the PubMed sentence length distribution and **(C)** is the PMC sentence distribution of each document and **(D)** is the PMC sentence length distribution.

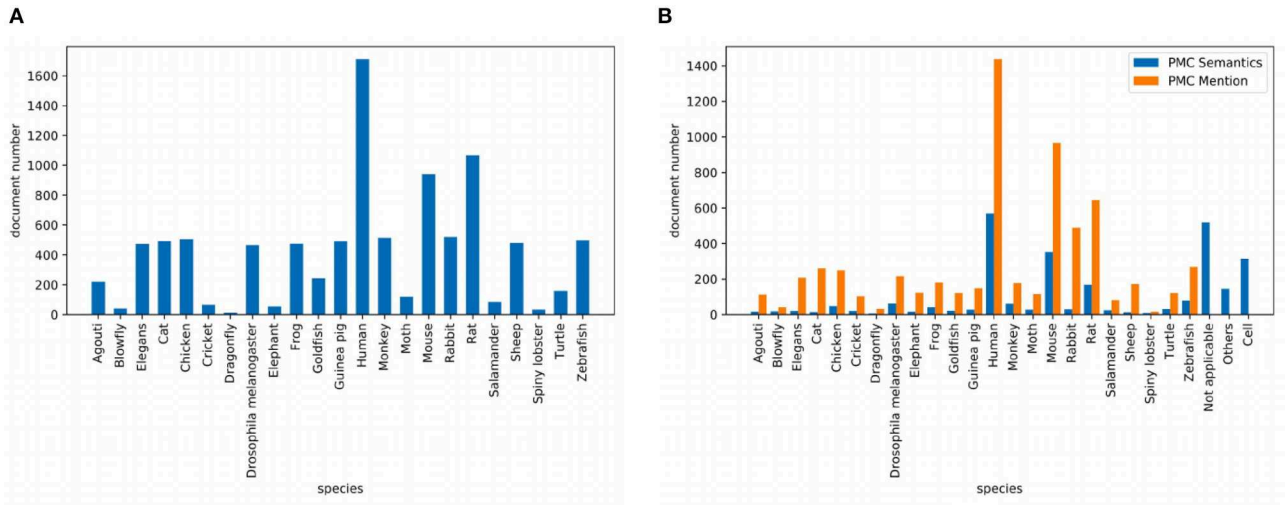

FIGURE 3 | Species distribution where **(A)** is the species distribution of the PubMed dataset and **(B)** is the species distribution of the PMC dataset.

TABLE 1 | The hyperparameter configuration.

Hyperparameters	Value
Character embedding	25
CNN kernel width	3
Encoder LSTM	100
Decoder LSTM	100
Dropout	0.5
Word embedding	GloVe.PubMed.200D
Epoch	100

vector representation. The main difference is the components they choose to encode the document.

The hierarchical CNN (H-CNN) and hierarchical LSTM (H-LSTM) use word- and sentence- level encoders to model the document structure, as shown in **Figures 4A,B** respectively. This is a hierarchical version of CNN and LSTM models.

The H-LSTM-ATT, also known as the hierarchical attention network (HAN) (Yang et al., 2016), adds an attention mechanism to the H-LSTM to extract informative words, as shown in **Figure 4C**. $c^{(w)}$ and $c^{(s)}$ are the word- and sentence- level context vectors respectively, and they can be trained jointly. To evaluate the influence of the LSTM layer, the H-MLP-ATT replaces the LSTM layer with a single layer neural network with the ReLU activation function, as shown in **Figure 4D**. This network can be seen as the H-CNN-ATT with the kernel size of $1 \times d$ where d is the vector dimension.

BERT (Devlin et al., 2019) is a pre-trained bidirectional transformer that has proven effective in various NLP tasks by fine-tuning the model. We use the representation of "[CLS]" to generate the document representation, as shown in **Figure 4E**. "[CLS]" stands for the representation of the class. Note that this model can only process up to 512 tokens.

4.3. Model Results

4.3.1. Results of the PubMed Dataset

Table 2 lists the results on the PubMed dataset. A first observation is that hierarchical models (H-LSTM and H-CNN) achieve similar results with the corresponding single-level models (LSTM and CNN) on the PubMed dataset. CNN models achieve higher results than the LSTM models in document classification. H-LSTM-ATT achieves better results than H-LSTM. This means the attention mechanism is important on this task. H-LSTM-ATT outperforms H-MLP-ATT, which means the LSTM layer encodes more context information of the sentence and the document. H-LSTM-ATT outperforms CNN, which further proves the importance of attention mechanism.

BERT achieves the highest result because fine-tuning this model allows it to adapt to a new target task. BERT's P/R/F1 per document are 0.7843/0.7994/0.7847. The drawback is that the model cannot encode the document structure and has the highest computation costs. Our SeqC model achieves comparable results. The P/R/F1 per document are 0.7588/0.7774/0.7612. **Figures 5A,B** show the class-aware results of SeqC and BERT respectively. The x- and y-axes denote the precision and recall respectively. The dotted lines are the contours of the F1. We observe that BERT achieves higher results on "Elegans, Moth, Elephant, Cat, Goldfish" classes. SeqC achieves higher results on "Agouti, Rat" classes. Other species achieve comparable prediction results on both models.

The dictionary-based method is most computationally efficient and easier to use, but it can be difficult to accomplish this task without mentioning species in the document. We evaluate this method in the case of restoring (+ Restore) the mentions of species in the literature. The dictionary-based method is a good choice when directly extracting the mentions of species. Restoring the mentions also significantly improves the SeqC model results. This is because the model will pay attention to the mentions of species.

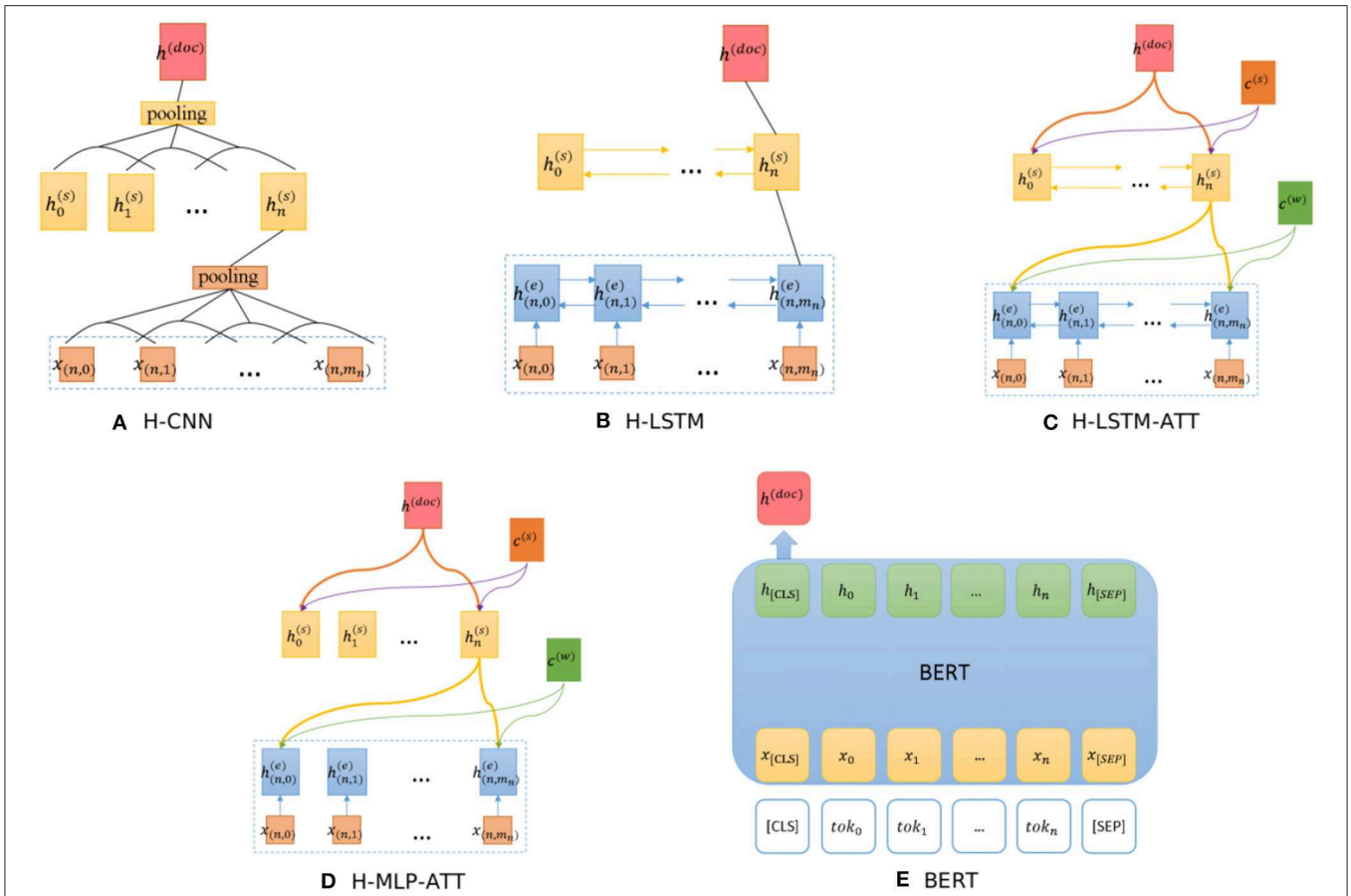

FIGURE 4 | Architectures of baseline models. **(A)** H-CNN, **(B)** H-LSTM, **(C)** H-LSTM-ATT, **(D)** H-MLP-ATT, **(E)** BERT.

TABLE 2 | Results of species classification on the PubMed dataset.

Algorithms	Hamming	Micro-F1	Macro-F1
LSTM (Zhang et al., 2015)	0.0302	79.01	73.33
CNN (Kim, 2014)	0.0247	82.84	81.13
H-LSTM	0.0292	79.86	74.09
H-CNN	0.0245	82.87	79.04
H-MLP-ATT	0.0275	81.47	80.53
H-LSTM-ATT	0.0228	84.35	84.24
BERT	**0.0204**	**86.20**	**86.03**
SeqC	0.0247	83.57	82.42
Dictionary + Restore	0.0029	97.98	**99.50**
SeqC + Restore	**0.0007**	**99.46**	99.39

Bold values represent the best results.

4.3.2. Results of the PMC Mention Dataset

Table 3 presents the results on the PMC dataset. We observe that CNN and LSTM models achieve comparable results on the PMC dataset. BERT achieves similar micro-F1 score with the H-LSTM-ATT model, but the macro-F1 score is higher than other models. This means that the overall performance of BERT is more balanced across classes. The simple SeqC model cannot predict the masked species well. When the SeqC model considers the discourse sections structure (+ Discourse), this method outperforms all baselines. The discourse sections structure denotes the section-level structure in the article's body. This model uses the word-discourse HAD, that is, considering the word-section level attention. This means the section-level information is important for extracting the SOIs of the article. This is because certain sections (e.g., the experiments section) can find research species more effectively. Longer documents contain more noise, which poses challenges for model prediction. The P/R/F1 per document of SeqC + Discourse are 0.7598/0.6901/0.7021. As shown in **Figures 6A,B**, we observe that BERT achieves higher results on "Human, Moth, Zebrafish" classes. Our model achieves higher results on "Mouse, Frog, Elephant, Drosophila melanogaster, Blowfly, Elegans, Monkey, Goldfish, Cricket, Guinea pig" classes. Other species achieve comparable prediction results on both models.

When we restore the mentions of species in the literature, the dictionary-based method outperforms other methods. Restoring mentions of species also significantly improves the results of our model when we extract species from the article's body.

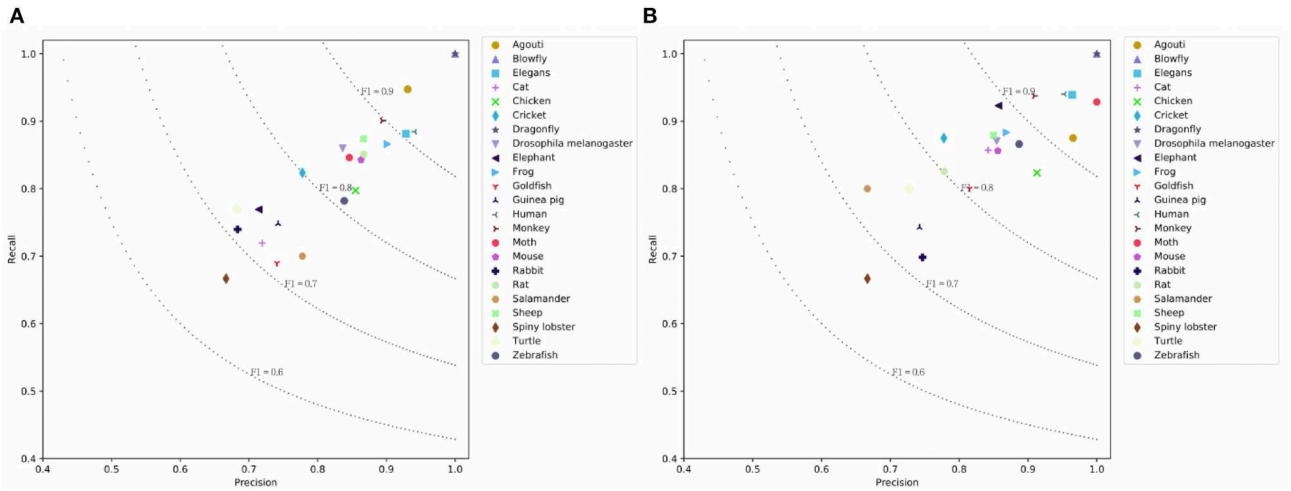

FIGURE 5 | Prediction results on the PubMed dataset where **(A)** is the prediction results using SeqC and **(B)** is the prediction results using BERT.

TABLE 3 | Results of species classification on the PMC Mention dataset.

Algorithms	Hamming	Micro-F1	Macro-F1
LSTM (Zhang et al., 2015)	0.0735	73.08	56.47
CNN (Kim, 2014)	0.0813	72.28	57.44
H-LSTM	0.0778	72.01	57.64
H-CNN	0.0760	72.78	57.05
H-MLP-ATT	0.0871	70.15	54.81
H-LSTM-ATT	0.0769	73.23	60.05
BERT	0.0767	73.93	63.02
SeqC	0.0889	70.26	55.91
SeqC + Discourse	**0.0655**	**76.81**	**64.41**
Dictionary + Restore	**0.0037**	**98.69**	**99.75**
SeqC + Discourse + Restore	0.0448	84.85	76.14

Bold values represent the best results.

4.3.3. Results of the PMC Semantics Dataset

Table 4 lists the results on the PMC Semantics dataset. We observe CNN models achieve higher results than the LSTM models. This means CNN units are good at capturing the internal semantics of documents. H-LSTM-ATT and H-CNN outperform the BERT. This means that the hierarchical modeling mechanism is good at capturing the document-level semantics. The simple SeqC does not perform well. The SeqC + Discourse achieves the highest performance. This means the section-level structure is more informative when modeling the article. This experiment proves our model is good at learning the semantic label of an article. As shown in **Figures 7A,B**, we observe that BERT achieves higher results on "Turtle, Salamander" classes. SeqC achieves higher results on "Spiny lobster, Zebrafish, Frog, Mouse, Rat, Goldfish, Cricket, Rabbit, Blowfly" classes. Other species achieve comparable prediction results on both models.

The PMC mention dataset is easier because the criteria of species mention are straightforward. The PMC Semantics dataset is more difficult because the annotation criteria are more complicated. The SeqC model can be more flexible to focus on different words for each species, which is helpful to let the model learn the annotation rule. This model frees researchers from tedious work and automatically classifies the literature. This experiment further proves the effectiveness of our models. The P/R/F1 per document is 0.8102/0.8/0.8006.

4.4. Analysis and Discussion

4.4.1. Ablation Study

To analyze the contributions and effects of different components, we perform ablation studies on the PubMed dataset, as shown in **Table 5**. The performance degrades by 1.83% micro-F1 without sentence-level attention (s-att). This is because the model cannot consider the sentence-level structure. The single-level attention only considers the word sequence, which assumes all sentences of a document are equally relevant for word selection. This setting limits the performance. When we remove the word-level attention (w-att), the performance drops by 2.02% micro-F1 and 4.28% macro-F1. This setting assumes that the contribution of all words in a sentence is the same, but the contribution of different sentences is different.

When we remove the HAD mechanism [s-att and word-level attention (w-att)], the performance drops by 3.62% micro-F1 and 4.61% macro-F1. This is because the model only uses the document vector to generate species and the decoder cannot attend to the document. When we remove the HAD mechanism and the decoder, the performance drops by 3.71% micro-F1 and 8.33% macro-F1. This is because the model becomes H-LSTM. The memory of a single document vector is limited.

4.4.2. Results of Different Species

It is instructive to analyze the prediction result of different species. **Figures 5A**, **6A**, **7A** visualize the class-aware prediction

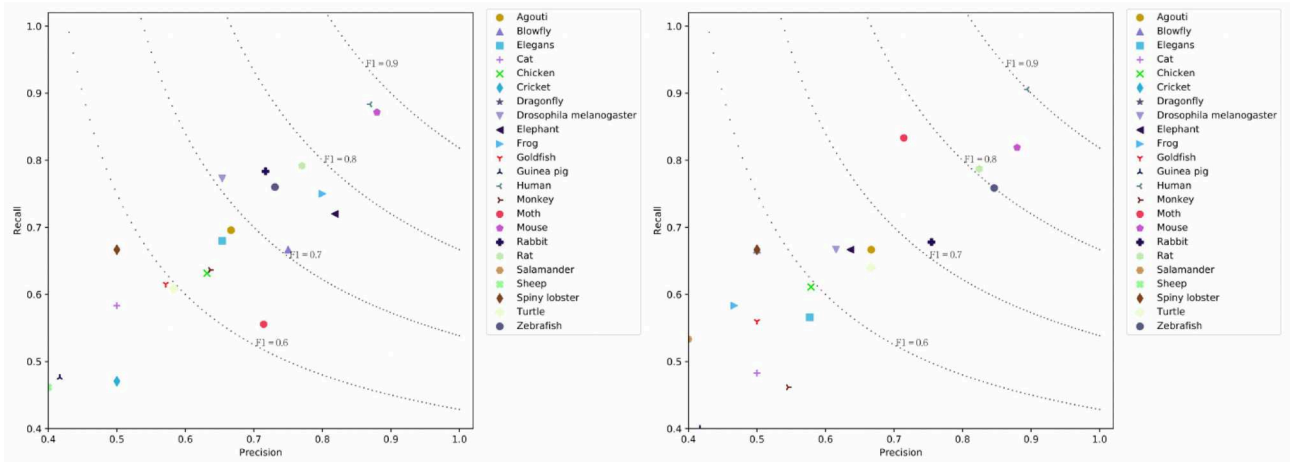

FIGURE 6 | Prediction results on the PMC Mention dataset where **(A)** is the prediction results using SeqC + Discourse and **(B)** is the prediction results using BERT.

TABLE 4 | Results of species classification on the PMC Semantics dataset.

Algorithms	Hamming	Micro-F1	Macro-F1
LSTM (Zhang et al., 2015)	0.0341	70.32	46.51
CNN (Kim, 2014)	0.0230	81.45	73.25
H-LSTM	0.0289	75.70	71.33
H-CNN	0.0213	82.17	72.05
H-MLP-ATT	0.0266	78.60	71.68
H-LSTM-ATT	0.0209	83.22	74.64
BERT	0.0230	81.57	72.12
SeqC	0.0246	80.91	70.34
SeqC + Discourse	**0.0203**	**84.03**	**74.75**
Dictionary + Restore	0.1270	42.50	35.38
SeqC + Discourse + Restore	**0.0189**	**85.41**	**79.03**

Bold values represent the best results.

results. The x- and y-axes represent the precision and recall respectively. The dotted lines denote the contours of the F1. For the PubMed dataset, we found "Dragonfly," "Blowfly," "Agouti," "Elegans," and "Human" are more easy to predict. The "Spiny lobster," "Rabbit," "Cat," and "Goldfish" are more problematic. For the PMC Mention dataset, we observe the "Human" and "Mouse" are easier to extract. The "Sheep," "Guinea pig," "Cricket," and "Cat" are more problematic. For the PMC Semantics dataset, we observe the "Elephant," "Spiny lobster," "Zebrafish" are easier to extract. The "Salamander" and "Others" are more problematic. We observe the prediction results are highly correlated to the class distribution.

As shown in **Figure 3B**, when we let experts annotate the corpus, the class imbalance problem has become more serious. This poses a challenge to the model. This phenomenon often occurs. Different versions of the annotated data have different class distributions. The forecasting of the results of the corpus annotation is important.

4.5. Species-Based Brain Cognitive Function, Brain Structure, and Protein Analysis

The hippocampus is a core brain region that is involved in many cognitive functions and brain diseases. The first part of **Table 6** lists part of the data and knowledge about brain diseases of different species extracted and analyzed using the proposed method. These diseases are considered related to the hippocampal study. This knowledge is also freely accessible on the Internet. We observe that some brain diseases are related to hippocampus, such as "Alpers' disease," "Anxiety," "Autism," "Brain edema," "Cerebral artery occlusion," "Lateral temporal epilepsy", etc. The research about "Lateral temporal epilepsy" is mainly conducted on "Human," "Rat," "Mouse", etc. Few studies are conducted based on the "Monkey," "Guinea pig," "Chicken," etc. Experiments with some innovative species could be instructive for gaining innovative insights into this disease. We can trace back to the scientific works based on the "Guinea pig," e.g., "The stimulation of 5-ht(1E) receptors and subsequent inhibition of adenylate cyclase activity in the DG suggests that 5-ht(1E) receptors may mediate regulation of hippocampal activity by 5-HT, making it a possible drug target for the treatment of neuropsychiatric disorders characterized by memory deficits (such as Alzheimer's disease) or as a target for the treatment of temporal lobe epilepsy (Klein and Teitler, 2012)."

The second part of **Table 6** lists part of the data and knowledge about cognitive functions of different species which are considered related to the hippocampal study. We observe that some cognitive functions are related to hippocampus, such as "Associative learning," "Aversion," "Acuity," "Concepts," "Decision making," "Olfactory," etc. Researchers prefer to conduct the researches for "Olfactory" on "Rat," "Mouse,"

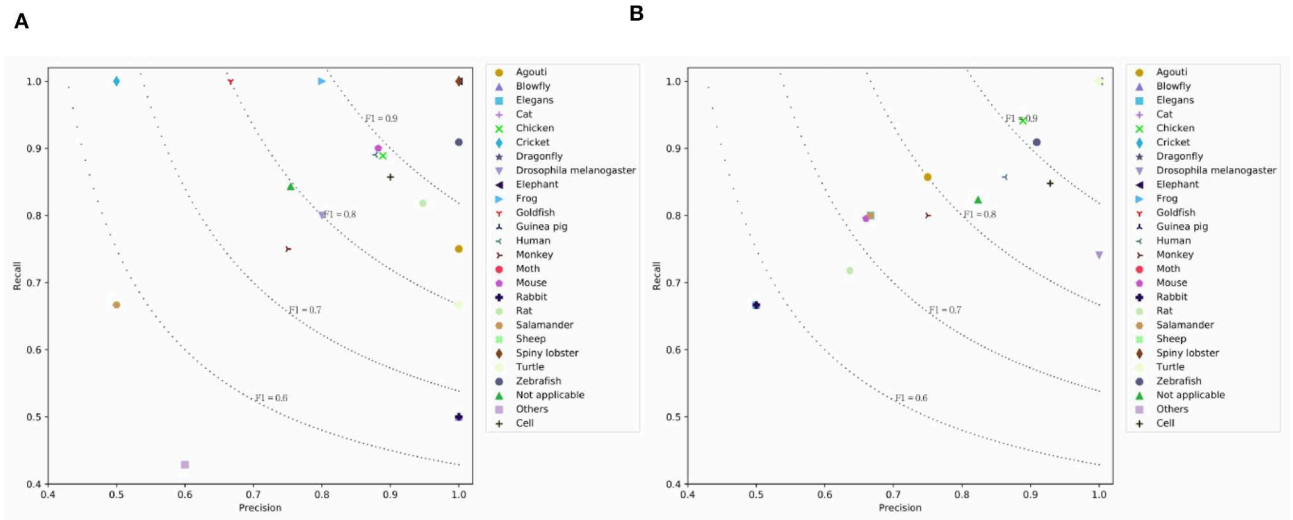

FIGURE 7 | Prediction results on the PMC Semantics dataset where **(A)** is the prediction results using SeqC + Discourse and **(B)** is the prediction results using BERT.

TABLE 5 | The ablation results on the PubMed dataset.

Model	Hamming	Micro-F1	Macro-F1
SeqC	0.0247	83.57	82.42
–s-att	0.0274	81.74	82.06
–w-att	0.0274	81.55	78.14
–HAD (s-att,w-att)	0.0300	79.95	77.81
–HAD (s-att,w-att), decoder	0.0292	79.86	74.09

"Human," etc. Few studies are conducted based on the "Monkey," "Sheep," "Guinea pig", etc. We found that research on monkeys' olfactory of smell may be relatively innovative. We can trace back to the scientific works based on the "Monkey," e.g., "Early developmental events involving the olfactory and limbic system start and conclude possibly slightly early in primates than rodents, and we find a comparable early conclusion of primate hippocampal neurogenesis (as assessed by the relative number of Ki67 cells) suggesting a plateau to low levels at approximately 2 years of age in humans (Charvet and Finlay, 2018)."

It can be found in the third part of **Table 6** that some proteins, such as "Acetylcholine esterase," "Adenosine deaminase," "Adenylate cyclase," "Aromatase," "Glutamine synthetase," "Nitric oxide synthase," etc., are related to the hippocampus. Researchers prefer to conduct the researches for "Nitric oxide synthase" on "Rat," "Mouse," "Human," etc. Few studies are conducted based on the "Guinea pig." We found that research on Guinea pig may be more instructive. For example, "Decreased nitric oxide synthase (NOS)-catalyzed formation of NO from L-arginine may be involved in ethanol teratogenesis involving the hippocampus (Gibson et al., 2000)."

4.6. Case Study

It is instructive to analyze how the attention mechanism extracts SOIs to predict species. We choose two abstracts (Zhou et al., 2017; Cho et al., 2019) to visualize the attention distribution, as shown in **Figures 8**, **9**. When the model predicts different species, it attends to different parts of the document. We restore the species names in the figure to better understand the samples. These species are marked with underlined stars.

For the first sample, this model first predicts "Human" by using the document representation. We observe this class is not mentioned in the abstract but is mentioned in the text so the "Human" can be assigned to this paper. This means our model can help infer more complete species. Some terms are potential topics in human-related research, e.g., "Huntington's disease," "Cognitive dysfunction," "huntingtin gene," "monogenetic disorder," etc. **Figure 8A** visualizes the attention distribution when predicting "Human." The attention distribution ("transgenic HD, N171-82Q, HD, neural, WT-NPCs, iPSCs") also contains information about the next species to be predicted, as this decoder sequentially models the correlation between species. When predicting "Mouse", the attention weight of "monogenetic, N171-82Q, neural progenitor, NPCs, pluripotent" increases and the weight of "iPSCs, WT-NPCs" decreases, as shown in **Figure 8B**. When predicting "EOS," token weights are distributed over all emphasized words and are most distracting, as shown in **Figure 8C**. This shows that the model attends to different words when predicting different species. The model also considers the correlation between labels and retains historical memory. However, this model misses "Monkey."

For the second sample, when predicting "Human," the model uses the document representation and attends to "neural, experimentation, nervous system, T-UCRs." When predicting "Monkey," the attention weights of "T-UCRs" and masked species words ("rhesus monkey") are increased. When predicting

TABLE 6 | Some examples of brain diseases, brain cognitive functions and proteins related to the brain region "hippocampus" in different species, where the number behind the species is the number of related studies.

Types	Examples	Species
Brain diseases	Alpers' disease	Cat (2), Chicken (1), Human (46), Mouse (21), Rabbit (8), Rat (62), Sheep (2)
	Anxiety	Cat (9), Human (119), Monkey (8), Mouse (206), Rat (241)
	Autism	Human (11), Monkey (1), Mouse (22), Rat (16)
	Brain edema	Cat (1), Human (2), Mouse (11), Rabbit (4), Rat (33)
	Cerebral artery occlusion	Cat (1), Human (2), Mouse (22), Rat (38)
	Lateral temporal epilepsy	Cat (1), Chicken (2), Guinea pig (3), Human (348), Monkey (3), Mouse (102), Rat (253), Zebrafish (1)
Brain cognitive functions	Associative learning	Human (19), Monkey (12), Mouse (24), Rabbit (5), Rat (34)
	Aversion	Cat (6), Human (52), Monkey (3), Mouse (96), Rabbit (6), Rat (356)
	Acuity	Human (1), Mouse (4), Rat (2)
	Concepts	Human(19), Monkey(2), Mouse(1), Rabbit(2), Rat(13)
	Decision making	Cat(1), Human(34), Mouse(6), Rat(26)
	Olfactory	Cat (5), Chicken (3), Frog (3), Guinea pig (7), Human (106), Monkey (7), Mouse (190), Rabbit (6), Rat (335), Sheep (8)
Proteins	Acetylcholine esterase	Cat (8), Guinea pig (8), Human (42), Monkey (6), Mouse (142), Rabbit (6), Rat (397)
	Adenosine deaminase	Human (1), Mouse (1), Rat (14)
	Adenylate cyclase	Cat (4), Chicken (1), Guinea pig (20), Human (18), Monkey (1), Mouse (31), Rabbit (1), Rat (150)
	Aromatase	Chicken (1), Human (15), Monkey (4), Mouse (30), Rat (42)
	Glutamine synthetase	Human (11), Mouse (10), Rabbit (2), Rat (41)
	Nitric oxide synthase	Guinea pig (9), Human (30), Mouse (89), Rat (240)

FIGURE 8 | Visualization of SOIs when the model predicts **(A)** Human **(B)** Mouse and **(C)** EOS where redness indicates attention and the stars below the text indicate the masked species.

FIGURE 9 | Visualization of SOIs when the model predicts **(A)** Human **(B)** Monkey **(C)** Mouse **(D)** Rat and **(E)** EOS where redness indicates attention and the stars below the text indicate the masked species.

"Mouse," the weights of "T-UCRs, nervous systems, neural stem" are increased. When predicting "Rat," the weights of "nervous systems, neural stem" are decreased. When predicting "EOS," token weights are most distracting.

5. CONCLUSION

We propose the SeqC framework to classify neuroscience literature for linking brain and neuroscience communities and devices on the Internet. This study facilitates knowledge transfer and real-time data analysis over the Internet. The advantages are that it is possible to visualize words that are receiving attention to make the model interpretable. Additionally, this could be used to infer more complete names of species. We use hierarchical encoders to model the document structure. We use a decoder with the HAD mechanism to extract SOIs for different species. To evaluate model performance, we create three datasets for species research of brain and neuroscience. We resolve the problem of species annotation and present two versions of annotation criteria (mention-based annotation and semantic-based annotation). Limitations are that labels should be provided before, and that a manual tagging is needed. However, the process is semi-automated and can be easily extended to a wider variety of species.

This paper uses deep learning models to resolve the problem of species classification for neuroscience literature. The proposed cognitive computing model resolves this problem primarily by attending to the SOIs of a document. This approach can help predict species in the neuroscience literature. Structured species knowledge can be used to inspire researchers to better understand the knowledge associations in brain and neuroscience. In the future, the limitations of manual labeling can be alleviated

by adding terms to the dictionary and using automatic model annotation. It seems promising to apply named entity recognition Zhu et al. (2019) models and attention mechanism to find more species names in the literature and perform open species extraction.

DATA AVAILABILITY STATEMENT

The datasets and codes generated for this study can be found in the Github https://github.com/sssgrowth/SPECIESEXPLORER.

AUTHOR CONTRIBUTIONS

YZ proposed the scientific question. HZ formalized the task, proposed the approach, annotated datasets and conducted the experiments. YZ contributed the domain terms, summarized the species, designed the ontology and upgraded the key insights of the model. HZ and YZ wrote the paper. DW collaborated to develop the models and data processing modules and annotate datasets. CH contributed the data annotation, upgraded the species annotation standard and discovered problems in the biological research process.

FUNDING

This study was supported by the Strategic Priority Research Program of Chinese Academy of Sciences (Grant No. XDB32070100) and the Beijing Municipality of Science and Technology (Grant No. Z181100001518006).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2020.00128/full#supplementary-material

REFERENCES

Ananiadou, S., and McNaught, J. (2006). *Text Mining for Biology and Biomedicine*. Norwood, MA: Citeseer.

Andersen, M. L. and Winter, L. M. (2017). Animal models in biological and biomedical research-experimental and ethical concerns. *An. Acad. Bras. Ciênc.* 91(Suppl. 1):e20170238. doi: 10.1590/0001-3765201720170238

Arunkumar, N., Mohammed, M. A., Mostafa, S. A., Ibrahim, D. A., Rodrigues, J. J., and de Albuquerque, V. H. C. (2020). Fully automatic model-based segmentation and classification approach for mri brain tumor using artificial neural networks. *Concurr. Comput. Pract. Exp.* 32:e4962. doi: 10.1002/cpe.4962

Ascoli, G. A., Donohue, D. E., and Halavi, M. (2007). Neuromorpho. org: a central resource for neuronal morphologies. *J. Neurosci.* 27, 9247–9251. doi: 10.1523/JNEUROSCI.2055-07.2007

Bada, M., Eckert, M., Evans, D., Garcia, K., Shipley, K., Sitnikov, D., et al. (2012). Concept annotation in the craft corpus. *BMC Bioinform.* 13:161. doi: 10.1186/1471-2105-13-161

Bahdanau, D., Cho, K., and Bengio, Y. (2015). "Neural machine translation by jointly learning to align and translate," in *Proceedings of ICLR* (San Diego, CA).

Bailey, J. (2006). *A Brief Overview of Chimpanzees and Aging Research*. Written for project R & R: Release and restitution for chimpanzees in US Laboratories (Ellensburg, WA).

Bebortta, S., Senapati, D., Rajput, N. K., Singh, A. K., Rathi, V. K., Pandey, H. M., et al. (2020). Evidence of power-law behavior in cognitive iot applications. *Neural Comput. Appl.* 32, 1–13. doi: 10.1007/s00521-020-04705-0

Bengio, S., Vinyals, O., Jaitly, N., and Shazeer, N. (2015). "Scheduled sampling for sequence prediction with recurrent neural networks," in *Proceedings of NIPS* (Montreal, QC).

Bodenreider, O. (2008). Biomedical ontologies in action: role in knowledge management, data integration and decision support. *Yearb. Med. Inform.* 17, 67–79. doi: 10.1055/s-0038-1638585

Charvet, C. J., and Finlay, B. L. (2018). Comparing adult hippocampal neurogenesis across species: translating time to predict the tempo in humans. *Front. Neurosci.* 12:706. doi: 10.3389/fnins.2018.00706

Che, W., Liu, Y., Wang, Y., Zheng, B., and Liu, T. (2018). "Towards better UD parsing: deep contextualized word embeddings, ensemble, and treebank concatenation," in *Proceedings of CoNLL 2018*, eds D. Zeman and J. Hajic (Brussels).

Chen, G., Ye, D., Xing, Z., Chen, J., and Cambria, E. (2017). "Ensemble application of convolutional and recurrent neural networks for multi-label text categorization," in *Proc. IJCNN* (Anchorage, AK: IEEE).

Cho, I. K., Hunter, C. E., Ye, S., Pongos, A. L., and Chan, A. W. S. (2019). Combination of stem cell and gene therapy ameliorates symptoms in huntington's disease mice. *npj Regener. Med.* 4:7. doi: 10.1038/s41536-019-0066-7

Cohan, A., Dernoncourt, F., Kim, D. S., Bui, T., Kim, S., Chang, W., et al. (2018). "A discourse-aware attention model for abstractive summarization of long documents," in *Proceedings of NAACL-HLT* (New Orleans, LA).

Cohen, K. B., and Demner-Fushman, D. (2014). *Biomedical Natural Language Processing*, Vol. 11. Amsterdam; Philadelphia, PA: John Benjamins Publishing Company.

Curtis, R. K., Orešič, M., and Vidal-Puig, A. (2005). Pathways to the analysis of microarray data. *Trends Biotechnol.* 23, 429–435. doi: 10.1016/j.tibtech.2005.05.011

De Albuquerque, V. H. C. Damaševičius, R., Garcia, N. M., Pinheiro, P. R., and Pedro Filho, P. R. (2017). Brain computer interface systems for neurorobotics: methods and applications. *Biomed. Res. Int.* 2017:2505493. doi: 10.1155/2017/2505493

Devlin, J., Chang, M., Lee, K., and Toutanova, K. (2019). 'BERT: pre-training of deep bidirectional transformers for language understanding," in *Proceedings of NAACL-HLT* (Minneapolis, MN).

Di Buccio, E., Li, Q., Melucci, M., and Tiwari, P. (2018). "Binary classification model inspired from quantum detection theory," in *Proceedings of the 2018 ACM SIGIR International Conference on Theory of Information Retrieval* (Tianjin), 187–190.

Dubitzky, W., Wolkenhauer, O., Yokota, H., and Cho, K.-H. (2013). *Encyclopedia of Systems Biology*. New York, NY: Springer Publishing Company, Incorporated.

Fan, R.-E., and Lin, C.-J. (2007). *A Study on Threshold Selection for Multi-Label Classification*. Department of Computer Science; National Taiwan University, Taiwan, 1–23.

Federhen, S. (2011). The ncbi taxonomy database. *Nucleic Acids Res.* 40, D136–D143. doi: 10.1093/nar/gkr1178

Funk, C., Baumgartner, W., Garcia, B., Roeder, C., Bada, M., Cohen, K. B., et al. (2014). Large-scale biomedical concept recognition: an evaluation of current automatic annotators and their parameters. *BMC Bioinform.* 15:59. doi: 10.1186/1471-2105-15-59

Gardner, D., Akil, H., Ascoli, G. A., Bowden, D. M., Bug, W., Donohue, D. E., et al. (2008). The neuroscience information framework: a data and knowledge environment for neuroscience. *Neuroinformatics* 6, 149–160. doi: 10.1007/s12021-008-9024-z

Gibson, M., Butters, N., Reynolds, J., and Brien, J. (2000). Effects of chronic prenatal ethanol exposure on locomotor activity, and hippocampal weight, neurons, and nitric oxide synthase activity of the young postnatal guinea pig. *Neurotoxicol. Teratol.* 22, 183–192. doi: 10.1016/S0892-0362(99)00074-4

Girshick, R. (2015). "Fast R-CNN," in *Proceedings of ICCV* (Santiago).

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of CVPR* (Columbus, OH).

Gochhayat, S. P., Kaliyar, P., Conti, M., Prasath, V., Gupta, D., and Khanna, A. (2019). Lisa: lightweight context-aware iot service architecture. *J. Clean. Prod.* 212, 1345–1356. doi: 10.1016/j.jclepro.2018.12.096

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). "Mask R-CNN," in *Proceedings of ICCV* (Venice).

Hemati, W., and Mehler, A. (2019). Crfvoter: gene and protein related object recognition using a conglomerate of crf-based tools. *J. Cheminform.* 11:21. doi: 10.1186/s13321-019-0343-x

Hersh, W. (2008). *Information Retrieval: A Health and Biomedical Perspective*. Berlin; Heidelberg: Springer Science & Business Media.

Hirschman, L., Yeh, A., Blaschke, C., and Valencia, A. (2005). Overview of biocreative: critical assessment of information extraction for biology. *BMC Bioinformatics* 6:S1. doi: 10.1186/1471-2105-6-S1-S1

Hoskins, W. T., and Pollard, H. P. (2005). Successful management of hamstring injuries in australian rules footballers: two case reports. *Chiropract. Osteopathy* 13:4. doi: 10.1186/1746-1340-13-4

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2008). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13. doi: 10.1093/nar/gkn923

Hunter, L., and Cohen, K. B. (2006). Biomedical language processing: what's beyond pubmed? *Mol. Cell* 21, 589–594. doi: 10.1016/j.molcel.2006.02.012

Imam, F. T., Larson, S., Grethe, J. S., Gupta, A., Bandrowski, A., and Martone, M. E. (2012). Development and use of ontologies inside the neuroscience information framework: a practical approach. *Front. Genet.* 3:111. doi: 10.3389/fgene.2012.00111

Jaiswal, A. K., Tiwari, P., Kumar, S., Gupta, D., Khanna, A., and Rodrigues, J. J. (2019). Identifying pneumonia in chest x-rays: a deep learning approach. *Measurement* 145, 511–518. doi: 10.1016/j.measurement.2019.05.076

Jensen, L. J., Saric, J., and Bork, P. (2006). Literature mining for the biologist: from information retrieval to biological discovery. *Nat. Rev. Genet.* 7:119. doi: 10.1038/nrg1768

Khatri, P., and Drăghici, S. (2005). Ontological analysis of gene expression data: current tools, limitations, and open problems. *Bioinformatics* 21, 3587–3595. doi: 10.1093/bioinformatics/bti565

Kim, Y. (2014). "Convolutional neural networks for sentence classification," in *Proceedings of EMNLP* (Doha).

Kingma, D., and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv preprint* arXiv:1412.6980.

Klein, M., and Teitler, M. (2012). Distribution of 5-ht1e receptors in the mammalian brain and cerebral vasculature: an immunohistochemical and pharmacological study. *Br. J. Pharmacol.* 166, 1290–1302. doi: 10.1111/j.1476-5381.2012.01868.x

Kumar, S., Tiwari, P., and Zymbler, M. (2019). Internet of things is a revolutionary approach for future technology enhancement: a review. *J. Big Data* 6:111. doi: 10.1186/s40537-019-0268-2

Kurata, G., Xiang, B., and Zhou, B. (2016). "Improved neural network-based multi-label classification with better initialization leveraging label co-occurrence," in *Proceedings of NAACL-HLT* (San Diego, CA).

Larson, S. D., and Martone, M. (2013). Neurolex. org: an online framework for neuroscience knowledge. *Front. Neuroinform.* 7:18. doi: 10.3389/fninf.2013.00018

Leach, D. R., Krummel, M. F., and Allison, J. P. (1996). Enhancement of antitumor immunity by ctla-4 blockade. *Science* 271, 1734–1736. doi: 10.1126/science.271.5256.1734

Li, J., Hu, R., Liu, X., Tiwari, P., Pandey, H. M., Chen, W., et al. (2019). A distant supervision method based on paradigmatic relations for learning word embeddings. *Neural Comput. Appl.* 31, 1–10. doi: 10.1007/s00521-019-04071-6

Liu, J., Chang, W.-C., Wu, Y., and Yang, Y. (2017). "Deep learning for extreme multi-label text classification," in *Proceedings of SIGIR* (Shinjuku).

Liu, X., Zeng, Y., Zhang, T., and Xu, B. (2016). Parallel brain simulator: a multi-scale and parallel brain-inspired neural network modeling and simulation platform. *Cogn. Comput.* 8, 967–981. doi: 10.1007/s12559-016-9411-y

Mallick, P. K., Ryu, S. H., Satapathy, S. K., Mishra, S., Nguyen, and Nhu, G. (2019). Brain mri image classification for cancer detection using deep wavelet autoencoder-based deep neural network. *IEEE Access* 7, 46278–46287. doi: 10.1109/ACCESS.2019.2902252

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., and McClosky, D. (2014). "The stanford corenlp natural language processing toolkit," in *Proceedings of ACL* (Baltimore, MD).

McNaughton, B., Barnes, C. A., and O'keefe, J. (1983). The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely-moving rats. *Exp. Brain Res.* 52, 41–49. doi: 10.1007/BF00237147

Micci, L., and Paiardini, M. (2016). Editorial overview: animal models for viral diseases. *Curr. Opin. Virol.* 19:9. doi: 10.1016/j.coviro.2016.08.014

Nam, J., Kim, J., Mencía, E. L., Gurevych, I., and Fürnkranz, J. (2014). "Large-scale multi-label text classification—revisiting neural networks," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (Nancy: Springer), 437–452.

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., et al. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proc. Natl. Acad. Sci. U.S.A.* 115, E5716–E5725. doi: 10.1073/pnas.1719367115

Poo, M.-M., Du, J.-L., Ip, N. Y., Xiong, Z.-Q., Xu, B., and Tan, T. (2016). China brain project: basic neuroscience, brain diseases, and brain-inspired computing. *Neuron* 92, 591–596. doi: 10.1016/j.neuron.2016.10.050

Qian, J., Tiwari, P., Gochhayat, S. P., and Pandey, H. M. (2020). A noble double dictionary based ecg compression technique for ioth. *IEEE Intern. Things J* 7:1. doi: 10.1109/JIOT.2020.2974678

Ren, S., He, K., Girshick, R., and Sun, J. (2015). "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proceedings of NIPS* (Montreal, QC), 91–99.

Riedel, S. (2005). "Edward jenner and the history of smallpox and vaccination," in *Baylor University Medical Center Proceedings*, Vol. 18 (Dallas, TX: Taylor & Francis), 21–25.

Sarmento, R. M., Vasconcelos, F. F., Filho, P. P. R., and de Albuquerque, V. H. C. (2020). An iot platform for the analysis of brain ct images based on parzen analysis. *Future Gener. Comput. Syst.* 105, 135–147. doi: 10.1016/j.future.2019.11.033

See, A., Liu, P. J., and Manning, C. D. (2017). "Get to the point: summarization with pointer-generator networks," in *Proceedings of ACL* (Vancouver, BC).

Silbert, L. C., Dodge, H. H., Lahna, D., Promjunyakul, N.-O., Austin, D., Mattek, N., et al. (2016). Less daily computer use is related to smaller hippocampal volumes in cognitively intact elderly. *J. Alzheimers Dis.* 52, 713–717. doi: 10.3233/JAD-160079

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., et al. (2007). The obo foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.* 25, 1251. doi: 10.1038/nbt1346

Sodhro, A. H., Fortino, G., Pirbhulal, S., Lodro, M. M., and Shah, M. A. (2017). "Energy efficiency in wireless body sensor networks," in *Networks of the Future: Architectures, Technologies, and Implementations*, eds M. Elkhodr, Q. F. Hassan, and S. Shahrestani (New York, NY: Chapman and Hall/CRC), 339.

Sodhro, A. H., Obaidat, M. S., Abbasi, Q. H., Pace, P., Pirbhulal, S., Fortino, G., et al. (2019a). Quality of service optimization in an iot-driven intelligent transportation system. *IEEE Wireless Commun.* 26, 10–17. doi: 10.1109/MWC.001.1900085

Sodhro, A. H., Pirbhulal, S., and de Albuquerque, V. H. C. (2019b). Artificial intelligence-driven mechanism for edge computing-based industrial applications. *IEEE Trans. Indus. Inform.* 15, 4235–4243. doi: 10.1109/TII.2019.2902878

Sodhro, A. H., Sangaiah, A. K., Sodhro, G. H., Lodro, M. M., Sekhari, A., Ouzrout, Y., et al. (2018). "Medical quality of service optimization over internet of multimedia things," in *Computational Intelligence for Multimedia Big Data on the Cloud with Engineering Applications*, eds A. K. Sangaiah, Z. Zhang, and M. Sheng (Cambridge, MA: Elsevier), 271–295.

Sunkin, S. M., Ng, L., Lau, C., Dolbeare, T., Gilbert, T. L., Thompson, C. L., et al. (2012). Allen brain atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Res.* 41, D996–D1008. doi: 10.1093/nar/gks1042

Tang, D., Qin, B., and Liu, T. (2015). "Document modeling with gated recurrent neural network for sentiment classification," in *Proceedings of EMNLP* (Lisbon).

Tiwari, P., and Melucci, M. (2018a). Multi-class classification model inspired by quantum detection theory. *arXiv preprint* arXiv:1810.04491.

Tiwari, P., and Melucci, M. (2018b). "Towards a quantum-inspired framework for binary classification," in *Proceedings of Information and Knowledge Management* (Torino), 1815–1818.

Tiwari, P., and Melucci, M. (2019a). "Binary classifier inspired by quantum theory," in *Proceedings of the AAAI Conference on Artificial Intelligence* (Honolulu, HI), Vol. 33, 10051–10052.

Tiwari, P., and Melucci, M. (2019b). Towards a quantum-inspired binary classifier. *IEEE Access* 7, 42354–42372. doi: 10.1109/ACCESS.2019.2904624

Vasconcelos, F. F., Sarmento, R. M., Filho, P. P. R., and de Albuquerque, V. H. C. (2020). Artificial intelligence techniques empowered edge-cloud architecture for brain ct image analysis. *Eng. Appl. Art. Intell.* 91:103585. doi: 10.1016/j.engappai.2020.103585

Venkatesan, R., and Er, M. J. (2014). "Multi-label classification method based on extreme learning machines," in *Proceedings of Control Automation Robotics & Vision (ICARCV)* (Singapore: IEEE), 619–624.

Wang, D., Tiwari, P., Garg, S., Zhu, H., and Bruza, P. (2020). Structural block driven enhanced convolutional neural representation for relation extraction. *Appl. Soft Comput.* 86:105913. doi: 10.1016/j.asoc.2019.105913

Wei, C.-H., Kao, H.-Y., and Lu, Z. (2015). Gnormplus: an integrative approach for tagging genes, gene families, and protein domains. *Biomed Res. Int.* 2015:918710. doi: 10.1155/2015/918710

Williams, R. J., and Zipser, D. (1989). A learning algorithm for continually running fully recurrent neural networks. *Neural Comput.* 1, 270–280. doi: 10.1162/neco.1989.1.2.270

Wiseman, S., and Rush, A. M. (2016). "Sequence-to-sequence learning as beam-search optimization," in *Proceedings of EMNLP* (Austin, TX).

Yang, P., Sun, X., Li, W., Ma, S., Wu, W., and Wang, H. (2018). "SGM: sequence generation model for multi-label classification," in *Proceedings of COLING* (Santa Fe, NM).

Yang, Z., Yang, D., Dyer, C., He, X., Smola, A. J., and Hovy, E. H. (2016). "Hierarchical attention networks for document classification," in *Proceedings of NAACL-HLT* (San Diego, CA).

Zeng, Y., Bi, W., Wang, Y., Tang, X., and Xu, B. (2014a). Automatic recovery of z-jumps for neuronal morphology reconstruction. *Front. Neuroinform.* 2014:2. doi: 10.3389/conf.fninf.2014.18.00002

Zeng, Y., Wang, D., Zhang, T., and Xu, B. (2014b). Linked neuron data (LND): a platform for integrating and semantically linking neuroscience data and knowledge. *Front. Neuroinform*. 2014:17. doi: 10.3389/conf.fninf.2014.18.00017

Zeng, Y., Wang, D., and Zhu, H. (2016). "User interests analysis and its application on the linked brain data platform," in *Proceedings of Brain Informatics (BI)*, eds G. A. Ascoli, M. Hawrylycz, H. H. Ali, D. Khazanchi, and Y. Shi (Omaha, NE).

Zeng, Y., Zhao, Y., Bai, J., and Xu, B. (2018). Toward robot self-consciousness (ii): brain-inspired robot bodily self model for self-recognition. *Cogn. Comput.* 10, 307–320. doi: 10.1007/s12559-017-9505-1

Zhang, M.-L., and Zhou, Z.-H. (2007). Ml-knn: a lazy learning approach to multi-label learning. *Pattern Recogn.* 40, 2038–2048. doi: 10.1016/j.patcog.2006.12.019

Zhang, X., Zhao, J., and LeCun, Y. (2015). "Character-level convolutional networks for text classification," in *Proceedings of NIPS* (Montreal, QC).

Zhao, F., Zeng, Y., Wang, G., Bai, J., and Xu, B. (2018). A brain-inspired decision making model based on top-down biasing of prefrontal cortex to basal ganglia and its application in autonomous uav explorations. *Cogn. Comput.* 10, 296–306. doi: 10.1007/s12559-017-9511-3

Zheng, B., Che, W., Guo, J., and Liu, T. (2016). "Chinese grammatical error diagnosis with long short-term memory networks," in *Proceedings of the 3rd Workshop on NLPTEA* (Osaka), 49–56.

Zhou, J., Wang, R., Zhang, J., Zhu, L., Liu, W., Lu, S., et al. (2017). Conserved expression of ultra-conserved noncoding rna in mammalian nervous system. *BBA Gene Regul. Mech.* 1860, 1159–1168. doi: 10.1007/s12559-017-9511-3

Zhu, H., Hu, W., and Zeng, Y. (2019). "Flexner: A flexible LSTM-CNN stack framework for named entity recognition," in *Proceedings of NLPCC*, eds J. Tang, M. Kan, D. Zhao, S. Li, and H. Zan (Dunhuang: Springer).

Zhu, H., Zeng, Y., and Wang, D. (2016a). "Brain knowledge engine," *Conference Abstract: Advances in Neuroinformatics* (Tokyo).

Zhu, H., Zeng, Y., Wang, D., and Xu, B. (2016b). "Brain knowledge graph analysis based on complex network theory," in *Proceedings of Brain Informatics (BI)* (Omaha, NE).

Zhu, H., Zeng, Y., Wang, D., and Xu, B. (2016c). "Relation inference and type identification based on brain knowledge graph," in *Proceedings of Brain Informatics (BI)* (Omaha, NE).

Zweigenbaum, P., Demner-Fushman, D., Yu, H., and Cohen, K. B. (2007). Frontiers of biomedical text mining: current progress. *Brief. Bioinform.* 8, 358–375. doi: 10.1093/bib/bbm045

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

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 31 July 2020
doi: 10.3389/fnhum.2020.00301

Anti-fatigue Performance in SSVEP-Based Visual Acuity Assessment: A Comparison of Six Stimulus Paradigms

Xiaowei Zheng[1], Guanghua Xu[1,2], Yubin Zhang[1], Renghao Liang[1], Kai Zhang[1], Yuhui Du[1], Jun Xie[1] and Sicong Zhang[1]*

[1] School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, China, [2] State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an, China

OPEN ACCESS

Edited by:
Marcin Wozniak,
Silesian University of Technology, Poland

Reviewed by:
Seong-Whan Lee,
Korea University, South Korea
Xiaogang Chen,
Peking Union Medical College, China
Sheng Ge,
Southeast University, China

***Correspondence:**
Guanghua Xu
ghxu@xjtu.edu.cn

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

Received: 22 May 2020
Accepted: 07 July 2020
Published: 31 July 2020

Citation:
Zheng X, Xu G, Zhang Y, Liang R, Zhang K, Du Y, Xie J and Zhang S (2020) Anti-fatigue Performance in SSVEP-Based Visual Acuity Assessment: A Comparison of Six Stimulus Paradigms. Front. Hum. Neurosci. 14:301. doi: 10.3389/fnhum.2020.00301

Purpose: The occurrence of mental fatigue when users stare at stimuli is a critical problem in the implementation of steady-state visual evoked potential (SSVEP)-based visual acuity assessment, which may weaken the SSVEP amplitude and signal-to-noise ratio (SNR) and subsequently affect the results of visual acuity assessment. This study aimed to explore the anti-fatigue performance of six stimulus paradigms (reverse vertical sinusoidal gratings, reverse horizontal sinusoidal gratings, reverse vertical square-wave gratings, brief-onset vertical sinusoidal gratings, reversal checkerboards, and oscillating expansion–contraction concentric rings) in SSVEP acuity assessment.

Methods: Based on four indices of $\alpha + \theta$ index, pupil diameter, National Aeronautics and Space Administration Task Load Index (NASA-TLX), and amplitude and SNR of SSVEPs, this study quantitatively evaluated mental fatigue in six SSVEP visual attention runs corresponding to six paradigms with 12 subjects.

Results: These indices of mental fatigue showed a good agreement. The results showed that the paradigm of motion expansion–contraction concentric rings had a superior anti-fatigue efficacy than the other five paradigms of conventional onset mode or pattern reversal mode during prolonged SSVEP experiment. The paradigm of brief-onset mode showed the lowest anti-fatigue efficacy, and the other paradigms of pattern reversal SSVEP paradigms showed a similar anti-fatigue efficacy, which was between motion expansion–contraction mode and onset mode.

Conclusion: This study recommended the paradigm of oscillating expansion–contraction concentric rings as the stimulation paradigm in SSVEP visual acuity because of its superior anti-fatigue efficacy.

Keywords: steady-state visual evoked potential, visual acuity, mental fatigue, anti-fatigue performance, stimulus paradigm

INTRODUCTION

Recently, there have been some research findings across a range of applications in vision science based on steady-state visual evoked potential (SSVEP) (Norcia et al., 2015; Odom et al., 2016; Zheng et al., 2019b). As an essential part of any ophthalmological or optometric examination, visual acuity is the most commonly measured visual function (Fahad et al., 2008). Within 40 years, the SSVEP technique has been used for measuring visual acuity in some studies, demonstrating that SSVEP provides an objective and quantitative method in visual acuity assessment, especially for infants or individuals with intellectual disabilities, hysteria, or malingering (Tyler et al., 1979; Norcia and Tyler, 1985a,b; Hemptinne et al., 2018).

There are some parameters, such as electrode placement, temporal frequency, stimulus area, and sweep duration, related to SSVEP visual acuity assessment, and some studies have given their recommended parameter settings (Yadav et al., 2009; Almoqbel et al., 2011; Hemptinne et al., 2018). As for stimulus paradigms used in SSVEP visual acuity assessment, previous studies have compared some performance, such as sensitive electrodes, harmonic components of SSVEP response, correlation, and agreement between objective SSVEP and subjective psychophysical visual acuity, of six paradigms (reverse vertical sinusoidal gratings, reverse horizontal sinusoidal gratings, reverse vertical square-wave gratings, brief-onset vertical sinusoidal gratings, reversal checkerboards, and oscillating expansion–contraction concentric rings) (Tobimatsu et al., 1993; Chen et al., 2019; Zheng et al., 2019a, 2020; Hamilton et al., 2020).

However, although SSVEP can be an objective method to assess visual acuity, mental fatigue caused by uncomfortable light twinkling and contrast changes of prolonged visual stimulus can decrease arousal level and attention, worsening the SSVEP signal quality and consequently degrading the practical performance (Lee et al., 2010; Zhu et al., 2010; Cao et al., 2014; Chen et al., 2015). Previous studies have indicated that the amplitude and the signal-to-noise ratio (SNR) are related to the mental fatigue of the subjects, with decreasing amplitude and SNR corresponding to the developing fatigue (Wu et al., 2010), which can affect the precision and the accuracy of SSVEP visual acuity results since the threshold determination criterion of SSVEP visual acuity is related to the amplitude and the SNR of electroencephalography (EEG) response (Fahad et al., 2008; Yadav et al., 2009).

To evaluate the mental fatigue of prolonged SSVEP task, previous studies have proved that EEGs in the α band (8–13 Hz) and the θ band (4–7 Hz) can be adapted to assess mental fatigue (Klimesch, 1999; Cao et al., 2014; Kathner et al., 2014; Xie et al., 2016). The θ activity is related to drowsiness, while the α waves appear during relaxed conditions, at decreased attention levels and in a drowsy but wakeful state (Klimesch, 1999). Increased fatigue level is often related to the global increase of EEG power in the α and the θ bands (Klimesch, 1999; Xie et al., 2016). Moreover, pupil diameter can also be an index to evaluate mental fatigue, and the increase of mental fatigue coincides with a decrease in pupil diameter (Hopstaken et al., 2015b; Koo et al., 2018). Besides that, the National Aeronautics and Space Administration Task Load Index (NASA-TLX) is also used as a subjective and quantitative estimation of mental fatigue (Hart and Staveland, 1988; Sampei et al., 2016).

On this basis, in this study, four indices, i.e., the EEG spectral powers of $\alpha + \theta$, SSVEP properties of amplitude and SNR, pupil diameters recorded by the eye tracker, and subjective NASA-TLX, were measured in six SSVEP visual attention runs corresponding to six previously mentioned types of paradigms to compare their anti-fatigue performance (Cao et al., 2014; Xie et al., 2016). We hypothesized that the reversal vertical sinusoidal gratings, reverse horizontal sinusoidal gratings, reverse vertical square-wave gratings, and reversal checkerboards would show a similar anti-fatigue performance since the stimulus mode and the pattern were similar. When staring at the brief-onset vertical sinusoidal gratings, the subjects would become more fatigued because of the constantly changing brightness of the onset and offset mode. As for the oscillating expansion–contraction concentric rings, since the overall brightness was uniform when evoking steady-state motion visual evoked potential (SSMVEP) (Xie et al., 2016; Zheng et al., 2019a), its anti-fatigue property would be better than that of other stimulus paradigms.

MATERIALS AND METHODS

Subjects

Twelve subjects (two females), aged between 21 and 25 years old and with normal or corrected normal visual acuity, participated in this experiment. They had no history of eye disease. All the subjects gave informed written consent following a protocol approved by the institutional review board of Xi'an Jiaotong University, conforming to the Declaration of Helsinki.

EEG Recordings

In this study, EEG signals were recorded from six occipital electrodes (PO3, PO4, POz, O1, O2, and Oz) according to the 10–20 system with a ground electrode, Fpz, placed on the forehead and a reference electrode, A1, placed on the left earlobe (Listed, 2006). The EEG signals were collected by a g.USBamp acquisition and processing system and an active electrode system g.GAMMAbox (g.tec, Schiedlberg, Austria) at a sampling rate of 1,200 Hz. Besides that, an online band-pass filter from 2 to 100 Hz was imposed to remove artifacts, and an offline notch filter between 48 and 52 Hz was applied to eliminate the power line interference.

Stimulus Designs

As shown in **Figure 1**, six stimulus paradigms (A: reverse vertical sinusoidal gratings, B: reverse horizontal sinusoidal gratings, C: reverse vertical square-wave gratings, D: brief-onset vertical sinusoidal gratings, E: reversal checkerboards, and F: oscillating expansion–contraction concentric rings) were introduced as six separate experimental runs (Zheng et al., 2020). As for each run, one stimulator was presented to the subjects at the center of a 24.5-in. LCD monitor (PG258Q, ASUS, Taipei, China) with a resolution of 1,920 × 1,080 pixels and a refresh rate of 240 Hz.

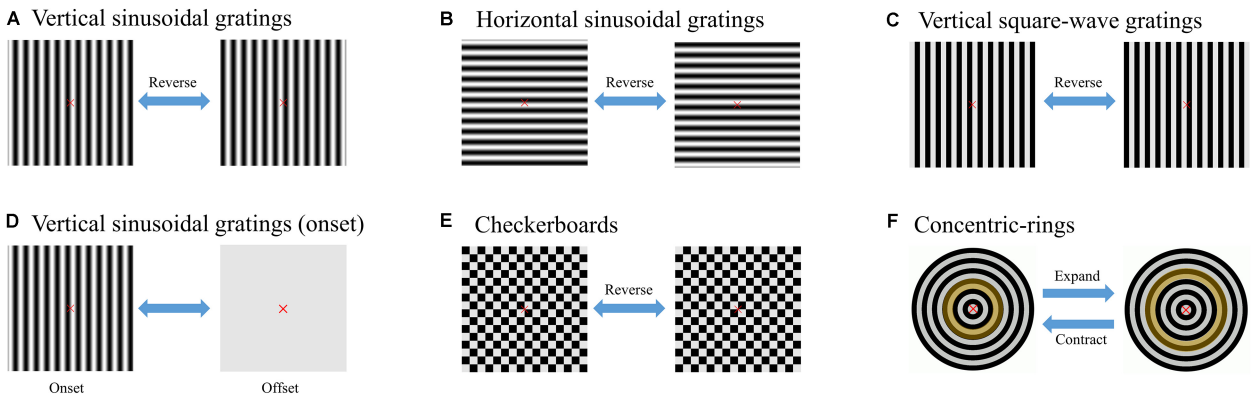

FIGURE 1 | Examples of six stimulus paradigms (Zheng et al., 2020). **(A)** Reverse vertical sinusoidal gratings. **(B)** Reverse horizontal sinusoidal gratings. **(C)** Reverse vertical square-wave gratings. **(D)** Brief-onset vertical sinusoidal gratings. **(E)** Reversal checkerboards. **(F)** Oscillating expansion–contraction concentric rings. The yellow ring indicates the shape shifting of the same zone during different processes in the paradigm of concentric rings.

The subjects were asked to sit 60 cm away from the monitor with the center at eye level. The visual angle of the stimulator was 4° with a diameter of 148 pixels, in accordance with the recommended visual angle parameter of previous studies (Almoqbel et al., 2011; Ng et al., 2012). The reversal or oscillating details of the six paradigms were the same as in our previous studies, with a contrast of 99.7%, and the duty cycle of paradigm D remained at 0.3 (Zheng et al., 2019a, 2020). According to previous studies (Almoqbel et al., 2011), the spatial frequency of three cycles per degree (cpd) corresponding to 1.0 logMAR optotype and temporal frequency of 7.5 Hz was assigned to all six stimulus paradigms. In the whole experiment, a spatially homogeneous white background with luminance of 208 cd/m^2 was displayed in pauses and around the stimulators. The stimulus paradigms were controlled by MATLAB (MathWorks, Natick, United States) with the Psychophysics Toolbox (Brainard, 1997).

Experimental Procedure

For each subject, six runs A, B, C, D, E, and F corresponding to six stimulus paradigms A, B, C, D, E, and F were carried out, respectively. An eye tracker (Tobii X2-30, Stockholm, Sweden) was used to monitor the subjects' eye movements and record their pupil diameter at a sample rate of 30 Hz. Each run consisted of 23 trials with three pre-experimental trials and 20 experimental trials. Each trial lasted 5 s, with an interval of 0.5 s between two trials. During the first three pre-experimental trials, to measure the baseline mental fatigue level from baseline pupil diameter and $\alpha + \theta$ band, the subjects stared at a black screen with only a red fixation cross at the position of the center stimulator, so there was no interference from the pupillary light reflexes of the eye to the environmental lighting (Hopstaken et al., 2015a). As for the other 20 trials, the stimulator was presented and the subjects were instructed to binocularly maintain attention on the center target stimulus throughout the experiment. The order of the six runs was random, and there was enough rest time for the subjects between two runs as long as the subjects wished. Additionally, a red fixation cross was presented at the center of the paradigms to aid fixation (Almoqbel et al., 2011). The whole experiment of each subject usually lasted for about 30–45 min, depending on the inter-run rest time governed by the subjects.

NASA-TLX

NASA-TLX, originally developed as a paper-and-pencil questionnaire by NASA Ames Research Center's Sandra Hart in the 1980s, has become a gold standard for measuring subjective workload across a wide range of applications (Hart and Staveland, 1988; Hart, 2006). Here we used NASA-TLX to assess mental fatigue subjectively as a psychological measurement. We assumed that workload deduced by the NASA-TLX represented the mental fatigue of the subjects (Sampei et al., 2016).

Firstly, the six defined sources of workload – mental demand (MD), physical demand (PD), temporal demand (TD), performance (PE), effort (EF), and frustration (FR) – were explained to the subjects.

The instruction was in Chinese as it is their native language. Then, after the subjects completed each run, they were asked to evaluate the six factors on a 0–100 scale. Next, after all the runs were completed, the subjects were asked to complete a pairwise comparison method of the six defined sources. The weights a, b, c, d, e, and f were assigned to each of the six workload sources from the pairwise comparison results, with weight integers ranging from 0 to 5, and their combinations were $C(6, 2) = 15$ (Sampei et al., 2016). Finally, the individual NASA-TLX of each run for each subject was derived from a weighted average of the ratings of these six factors:

$$\mathrm{NASA-TLX} = \frac{a * \mathrm{MD} + b * \mathrm{PD} + c * \mathrm{TD} + d * \mathrm{PE} + e * \mathrm{EF} + f * \mathrm{FR}}{15}. \quad (1)$$

Signal Processing of the EEG Data

Canonical Correlation Analysis

Canonical correlation analysis (CCA) is a non-parametric multivariable method used to reflect the overall linear correlation between two groups of variables, and it is also used in the analysis

of SSVEPs (Lin et al., 2007; Bin et al., 2009; Yan et al., 2018). In our study, it was also used to describe the correlations between the multi-channel SSVEP signals X and reference signal Y_i. X is the six-electrode channel signal in each trial. The reference signal Y_i composed of sine and cosine pairs is constructed at the reference frequency f_i ($i = 1, 2, \ldots, N$):

$$Y_i = \begin{pmatrix} \sin\left(2\pi f_i t\right) \\ \cos(2\pi f_i t) \end{pmatrix}, \quad t = \frac{1}{F_s}, \ldots, \frac{S}{F_s}. \tag{2}$$

where F_s is the sampling rate, and S is the sample point. Here the reference frequency f_i is set to 1.0, 1.1, . . ., 35.0 Hz (i.e., $N = 341$).

The linear transformations of X and Y_i are $x = {w_x}^T X$ and $y_i = {w_{yi}}^T Y_i$, respectively, and the maximum correlation coefficient value ρ_i between X and Y_i can be calculated by the CCA method as:

$$\rho_i = \max_{w_x, w_y} \frac{E[w_x^T X Y_i^T w_{yi}]}{\sqrt{E\left[w_x^T X X^T w_x\right] E[w_{yi}^T Y_i Y_i^T w_{yi}]}}. \tag{3}$$

where E denotes the symbol of the expected value in statistics, and the superscript T indicates the transposed matrix. The maximum correlation coefficient value ρ_i, which represents the maximum correlation between X and Y_i, can be considered as the response to the stimulus paradigm of SSVEPs at the reference frequency f_i ($f_1, f_2 \ldots, f_N$). Therefore, all the ρ_i and their corresponding frequency f_i can be plotted as a CCA spectrum. The ρ_i at the stimulus frequency of 7.5 Hz was regarded as the SSVEP amplitude.

Signal-to-Noise Ratio

Signal-to-noise ratio (SNR) refers to the ratio of signal to noise in a device or system. In our study, the SNR was defined as the ratio of the square of the CCA coefficient at the stimulus frequency of 7.5 Hz to the mean value of the square of the n adjacent points on the CCA spectrum:

$$\mathrm{SNR} = \frac{z(f)^2}{\frac{1}{n} * \sum_{k=1}^{\frac{n}{2}} \left[z\left(f + c * k\right)^2 + z\left(f - c * k\right)^2\right]}. \tag{4}$$

where n is set to 10, and f is 7.5 Hz. $z(f)$ is the CCA coefficient of the stimulus frequency f on the CCA spectrum. Then, c is the scale value of abscissa on the CCA spectrum, which is set to 0.1.

EEG Spectral Powers of α + θ Band

Common average reference (CAR) fusion is a commonly used EEG spatial filtering method performed by subtracting the mean of all electrode signals from the selected electrode signals to enhance the SNR of the selected electrode signals (Friman et al., 2007; Yan et al., 2019). In this study, we chose Oz electrode in spectral analysis, so the time domain EEG signal V_i to be analyzed can be expressed as:

$$V_i = V_{Oz} - \frac{1}{6}\sum_{j=1}^{6} V_j, \tag{5}$$

where V_j is the EEG signal from six electrode channels (PO3, PO4, POz, O1, O2, and Oz).

As for the processing of the EEG signals, firstly, a band-pass filter of 3–45 Hz was carried out to remove low-frequency drift and high-frequency interference. Then, CAR fusion was used for spatial filtering in each trial. Next, the Welch power spectrum density (PSD) in bins of 0.1 Hz was used for spectral analysis to obtain the EEG spectral powers of the α band of 8–13 Hz and the θ band of 4–7 Hz. Finally, the sum value of the PSD amplitude in the frequency band on the Welch power spectrum was defined as the EEG band power indices of the frequency bands of α and θ (Cao et al., 2014). Hence, the EEG combined index (α + θ) in each trial was obtained.

As the stimulation time and the experimental trial increased, the subject could get more fatigued. Hence, the mean values and SD of each index in the 1–5, 6–10, 11–15, and 16–20 experimental trials of each run were used to represent the corresponding four fatigue levels (i.e., level 1, level 2, level 3, and level 4), respectively (Xie et al., 2016). Fatigue level 4 represented the most fatigued state, while level 1 represented the least fatigued state.

Statistical Analysis

Statistical analyses were carried out using SPSS 22.0 (IBM, Armonk, United States). One-way or two-way repeated-measures analysis of variance (ANOVA) with a significance of $P < 0.05$ was employed to evaluate the significance of changes in the four indices of α + θ index, SSVEP amplitude and SNR, pupil diameter index, and NASA-TLX index of six paradigms at two fatigue levels, i.e., fatigue level 1 and level 4. The *post-hoc* analysis with Bonferroni correction method for multiple comparisons was also used when necessary. Besides that, we used equal signs and inequality signs to visualize the anti-fatigue performance among the six paradigms based on each index.

RESULTS

Pre-experimental Trials

As for each subject, we assumed that the initial mental fatigue was the same at the beginning of each run since there was enough rest time between two runs and the order of the presentation of the six runs was random. Here, to verify this assumption, we estimated the difference of initial baseline mental fatigue among six runs corresponding to six paradigms. As the first three pre-experimental trials of each run were presented with a black background, the mean fatigue level of the first three pre-experimental trials can be regarded as the initial mental fatigue for each run. One-way repeated-measures ANOVA was used to analyze the difference in pupil diameter and α + θ band of the first three pre-experimental trials for each paradigm. As shown in **Figure 2**, there was no significant difference both in the pupil diameter and the α + θ band of the first three pre-experimental trials for each paradigm [$F_{(5,55)} = 0.687$, $P = 0.635$ for pupil diameter; Greenhouse–Geisser correction: $F_{(3.238,35.623)} = 0.774$, $P = 0.525$ for the α + θ band], demonstrating that our assumption that the initial mental fatigue was the same at the beginning of each run was credible.

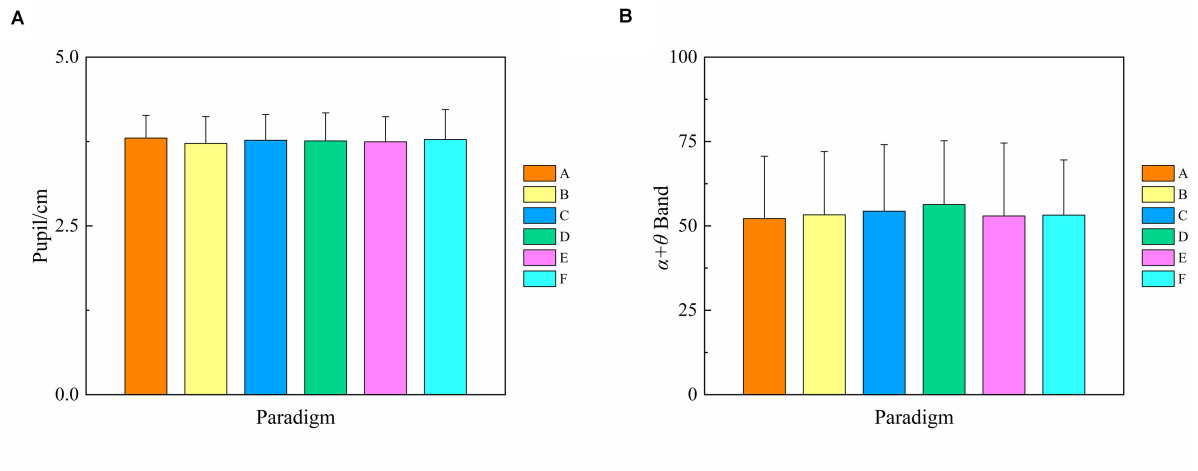

FIGURE 2 | Comparison of the mean values and SD of the pupil diameter and α + θ band index of the first three pre-experimental trials for each paradigm over 12 subjects. Statistics were assessed by one-way repeated-measures ANOVA. **(A)** Pupil diameter index. **(B)** α + θ band index.

Comparison of NASA-TLX

For the convenience of data analysis, the NASA-TLX of psychological measurement of mental fatigue was normalized per subject by his/her maximal value of six workload sources. The mean values and SD of normalized NASA-TLX for six stimulation paradigms over 12 subjects are shown in **Figure 3**. One-way repeated-measures ANOVA revealed that there was a significant difference in NASA-TLX among six paradigms [$F_{(5,55)} = 0.074$, $P = 0.044$]. As the corresponding Bonferroni *post-hoc* analysis shows in **Table 1**, there was no difference in the mean values of NASA-TLX among paradigms A, B, C, and E, demonstrating that these pattern reversal paradigms had a similar stimulus intensity for the human eyes. Paradigm D had the highest mean value of NASA-TLX than the other paradigms ($P < 0.05$, respectively), demonstrating that visual stimulation of onset mode had a high stimulus intensity, which may be due to the repetitive attentional demands of continuous flicker and contrast change (Xie et al., 2017). Except for paradigms C and E with a slightly but non-significantly higher NASA-TLX than paradigm F, there was a significant difference in the mean values of NASA-TLX between paradigms F and A, B, and D ($P < 0.05$, respectively), showing that paradigm F had the lowest value of NASA-TLX, in favor of that motion expansion–contraction SSMVEP stimulation which exhibited a superior anti-fatigue efficacy over the conventional flickering or pattern reversal SSVEP stimulation (Xie et al., 2016). Hence, the anti-fatigue performance of six paradigms based on NASA-TLX was as follows: F > E = A = B = C > D.

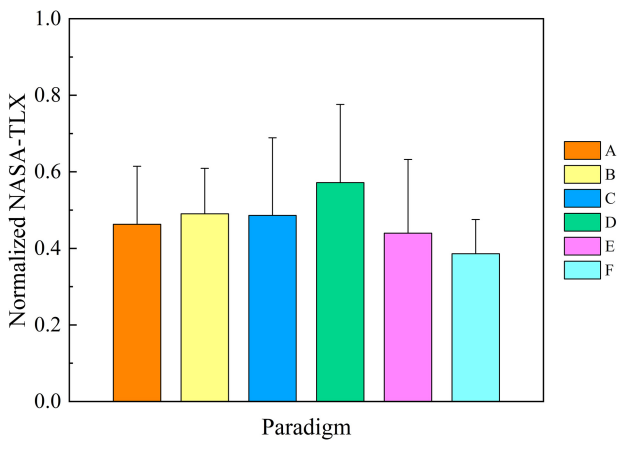

FIGURE 3 | Comparison of the mean values and SD of normalized National Aeronautics and Space Administration Task Load Index for six stimulus paradigms over 12 subjects. Statistics were assessed by one-way repeated-measures ANOVA.

Comparison of SSVEP Amplitude and SNR

To compare the changes in amplitude and SNR, the mean values and SD of the SSVEP amplitude and SNR summed over the stimulus frequency of 7.5 Hz in the 1–5 and 16–20 trials of each run were grouped to represent fatigue levels 1 and 4, respectively, as shown in **Figure 4**. Two-way repeated-measures ANOVA revealed that the interaction of two factors

TABLE 1 | Bonferroni *post-hoc* analysis of National Aeronautics and Space Administration Task Load Index among six stimulus paradigms.

Paradigm	B	C	D	E	F
A	$P = 0.273$	$P = 0.436$	$P = 0.010$*	$P = 0.374$	$P = 0.038$*
B	–	$P = 0.828$	$P = 0.016$*	$P = 0.183$	$P < 0.001$***
C	–	–	$P = 0.008$**	$P = 0.058$	$P = 0.054$
D	–	–	–	$P = 0.005$**	$P = 0.002$**
E	–	–	–	–	$P = 0.302$
F	–	–	–	–	–

****P < 0.001; **P < 0.01; *P < 0.05.*

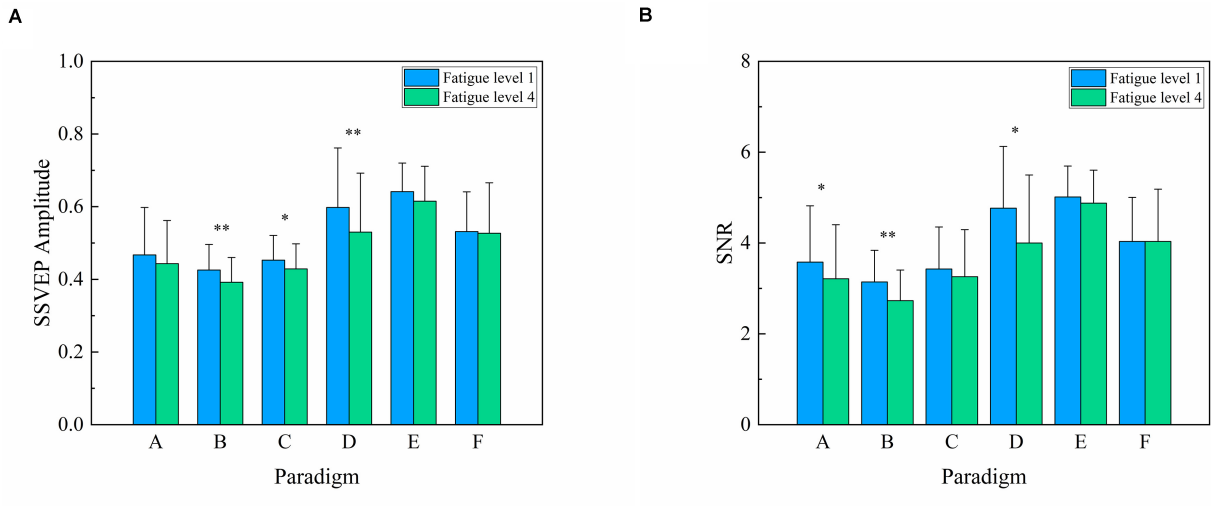

FIGURE 4 | Comparison of the mean values and SD of steady-state visual evoked potential amplitude and signal-to-noise ratio (SNR) between fatigue level 1 and fatigue level 4 for each paradigm over 12 subjects. **(A)** Amplitude. **(B)** SNR. Statistics were assessed by one-way repeated-measures ANOVA. $^{**}P < 0.01$; $^{*}P < 0.05$.

of "stimulus paradigm" and "fatigue level" yielded significance in SSVEP amplitude [$F_{(5,55)} = 2.955$, $P = 0.020$] and SNR [$F_{(5,55)} = 2.695$, $P = 0.030$]. Subsequently, one-way repeated-measures ANOVA found a significant difference in SSVEP amplitude among six paradigms at fatigue level 1 [Greenhouse–Geisser $F_{(2.510,27.609)} = 14.116$, $P < 0.001$] and fatigue level 4 [Greenhouse–Geisser $F_{(2.637,29.006)} = 11.019$, $P < 0.001$]. As the corresponding Bonferroni *post-hoc* analysis of SSVEP amplitude among six paradigms at fatigue level 1 and level 4 shows in **Tables 2**, **3**, the SSVEP amplitude induced by different stimulus paradigms was different. Paradigms E and D induced the highest amplitude, and paradigms A, B, and C induced the lowest amplitude, with paradigm F in between.

One-way repeated-measures ANOVA was also used to analyze the difference in SSVEP amplitude for paradigms A, B, C, D, E, and F between fatigue level 1 and level 4 and found a significant decrease in paradigms B, C, and D [$F_{(1,11)} = 12.201$, $P = 0.005$ for paradigm B; $F_{(1,11)} = 5.047$, $P = 0.046$ for paradigm C; $F_{(1,11)} = 13.749$, $P = 0.003$ for paradigm D]. The same trend of decrease was also found in paradigms A and E without statistical significance [$F_{(1,11)} = 2.550$, $P = 0.139$ for paradigm A; $F_{(1,11)} = 4.353$, $P = 0.061$ for paradigm E]. However, there was no obvious change in amplitude for paradigm F [$F_{(1,11)} = 0.083$, $P = 0.779$].

A similar significant difference of SNR results can also be found among six paradigms at fatigue level 1 [Greenhouse–Geisser $F_{(2.427,26.696)} = 11.949$, $P < 0.001$] and fatigue level 4 [Greenhouse–Geisser $F_{(2.688,29.569)} = 10.594$, $P < 0.001$] by one-way repeated-measures ANOVA. The corresponding Bonferroni *post-hoc* analysis of SSVEP SNR among six paradigms at fatigue level 1 and level 4 is shown in **Tables 4**, **5**, revealing that paradigms E and D had the highest SNR and paradigms A, B, and C had the lowest SNR, with paradigm F in between.

One-way repeated-measures ANOVA found that SNR had a similar decrease for paradigms A, B, and D between fatigue level 1 and level 4 [$F_{(1,11)} = 6.175$, $P = 0.030$ for paradigm A; $F_{(1,11)} = 12.471$, $P = 0.005$ for paradigm B; $F_{(1,11)} = 8.584$, $P = 0.014$ for paradigm D]. The same but non-significant trend of decrease was also found in paradigm C [$F_{(1,11)} = 1.103$, $P = 0.316$] and paradigm E [$F_{(1,11)} = 1.341$, $P = 0.271$]. However, there was no obvious change in SNR for paradigm F [$F_{(1,11)} < 0.001$, $P = 0.999$].

TABLE 2 | Bonferroni *post-hoc* analysis of steady-state visual evoked potential amplitude among six paradigms at fatigue level 1.

Paradigm	B	C	D	E	F
A	$P = 1.000$	$P = 1.000$	$P = 0.228$	$P = 0.204$	$P = 1.000$
B	–	$P = 1.000$	$P = 0.022^{*}$	$P < 0.001^{***}$	$P = 0.006^{**}$
C	–	–	$P = 0.116$	$P < 0.001^{***}$	$P = 0.086$
D	–	–	–	$P = 1.000$	$P = 1.000$
E	–	–	–	–	$P = 0.017^{*}$
F	–	–	–	–	–

$^{***}P < 0.001$; $^{**}P < 0.01$; $^{*}P < 0.05$.

TABLE 3 | Bonferroni *post-hoc* analysis of steady-state visual evoked potential amplitude among six paradigms at fatigue level 4.

Paradigm	B	C	D	E	F
A	$P = 0.695$	$P = 1.000$	$P = 1.000$	$P = 0.018^{*}$	$P = 1.000$
B	–	$P = 0.343$	$P = 0.109$	$P < 0.001^{***}$	$P = 0.024^{*}$
C	–	–	$P = 0.459$	$P < 0.001^{***}$	$P = 0.094$
D	–	–	–	$P = 1.000$	$P = 1.000$
E	–	–	–	–	$P = 0.533$
F	–	–	–	–	–

$^{***}P < 0.001$; $^{**}P < 0.01$; $^{*}P < 0.05$.

TABLE 4 | Bonferroni *post-hoc* analysis of steady-state visual evoked potential signal-to-noise ratio among six paradigms at fatigue level 1.

Paradigm	B	C	D	E	F
A	$P = 1.000$	$P = 1.000$	$P = 0.311$	$P = 0.025$	$P = 1.000$
B	–	$P = 0.654$	$P = 0.043$*	$P < 0.001$***	$P = 0.060$
C	–	–	$P = 0.198$	$P < 0.001$***	$P = 0.298$
D	–	–	–	$P = 1.000$	$P = 0.501$
E	–	–	–	–	$P = 0.015$*
F	–	–	–	–	–

****P < 0.001; **P < 0.01; *P < 0.05.*

TABLE 5 | Bonferroni *post-hoc* analysis of steady-state visual evoked potential signal-to-noise ratio among six paradigms at fatigue level 4.

Paradigm	B	C	D	E	F
A	$P = 1.000$	$P = 1.000$	$P = 1.000$	$P = 0.008$	$P = 0.734$
B	–	$P = 0.576$	$P = 0.274$	$P < 0.001$***	$P = 0.044$*
C	–	–	$P = 1.000$	$P < 0.001$***	$P = 0.247$
D	–	–	–	$P = 1.000$	$P = 1.000$
E	–	–	–	–	$P = 0.134$
F	–	–	–	–	–

****P < 0.001; **P < 0.01; *P < 0.05.*

This implied that the factor of the types of stimulus paradigm had a significant influence on the SSVEP response during prolonged usage. Both SSVEP amplitude and SNR between fatigue level 1 and fatigue level 4 had a downtrend for paradigms A, B, C, D, and E, but paradigm F did not present a significant change in SSVEP amplitude and SNR between fatigue level 1 and level 4. These results were also in line with previous studies such that paradigm F of motion SSMVEP stimulation exhibited a superior anti-fatigue efficacy than conventional flickering or pattern reversal SSVEP stimulation during prolonged SSVEP visual acuity assessment (Xie et al., 2016). Hence, the anti-fatigue performance of six paradigms based on SSVEP amplitude and SNR was as follows: F > E > A = C = B ≥ D.

Comparison of Pupil Diameter Index

For the convenience of data analysis, the pupil diameter corresponding to each paradigm was normalized for each subject by his/her respective baseline pupil diameter of the first three pre-experimental trials of each paradigm. **Figure 5A** presents the normalized pupil diameter index between fatigue level 1 and fatigue level 4 for six stimulus paradigms over 12 subjects. Two-way repeated-measures ANOVA revealed that the interaction of two factors of "stimulus paradigm" and "fatigue level" yielded significance in normalized pupil diameter index [$F_{(5,55)} = 2.727$, $P = 0.029$]. Subsequently, one-way repeated-measures ANOVA found no significant difference in pupil diameter index among six paradigms at fatigue level 1 [$F_{(5,55)} = 1.796$, $P = 0.129$] and fatigue level 4 [$F_{(5,55)} = 0.170$, $P = 0.973$].

One-way repeated-measures ANOVA was also used to analyze the difference in pupil diameter index for all six paradigms between fatigue level 1 and level 4 and found a significant decrease in paradigms A, B, C, D, and E [$F_{(1,11)} = 18.291$, $P = 0.001$ for paradigm A; $F_{(1,11)} = 15.803$, $P = 0.002$ for paradigm B; $F_{(1,11)} = 15.226$, $P = 0.002$ for paradigm C; $F_{(1,11)} = 8.134$, $P = 0.016$ for paradigm D; $F_{(1,11)} = 13.177$, $P = 0.004$ for paradigm E). However, there was no obvious change in pupil diameter index for paradigm F [$F_{(1,11)} = 0.091$, $P = 0.769$]. This also revealed that paradigm F had better anti-fatigue efficacy than the other five paradigms. Hence, the anti-fatigue performance of the six paradigms based on the pupil diameter index was as follows: F > E = A = B = C = D.

Comparison of α + θ Index

The α + θ band corresponding to each paradigm was normalized for each subject by his/her respective baseline α + θ band of the first three pre-experimental trials of each paradigm. **Figure 5B** presents the normalized α + θ index between fatigue level 1 and

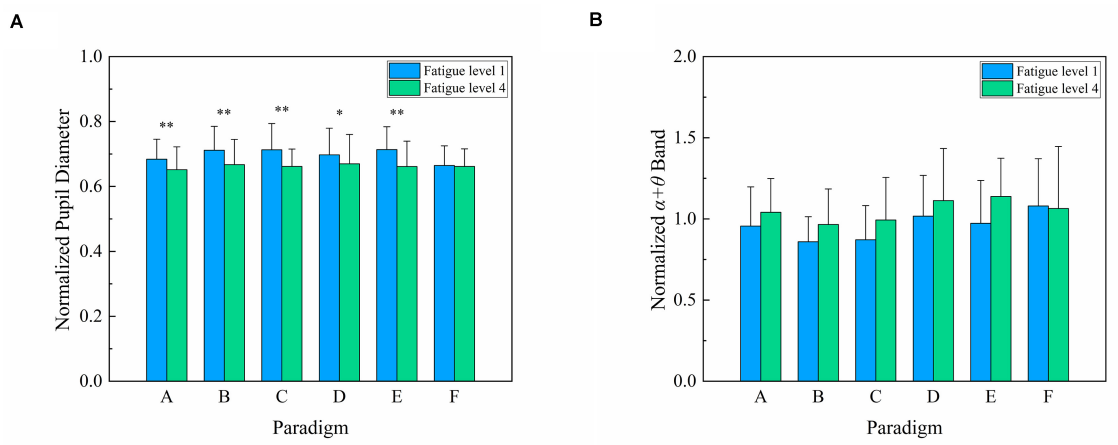

FIGURE 5 | Comparison of the mean values and SD of normalized pupil diameter index and normalized α + θ index between fatigue level 1 and fatigue level 4 for six stimulus paradigms over 12 subjects. **(A)** Normalized pupil diameter index. Statistics were assessed by repeated-measures ANOVA. **$P < 0.01$; *$P < 0.05$. **(B)** Normalized α + θ index.

TABLE 6 | Bonferroni *post-hoc* analysis of the α + θ index among six paradigms.

Paradigm	B	C	D	E	F
A	$P = 1.000$	$P = 1.000$	$P = 0.362$	$P = 0.010$	$P = 1.000$
B	–	$P = 0.479$	$P = 0.072$	$P < 0.001$***	$P = 0.035$*
C	–	–	$P = 0.278$	$P < 0.001$***	$P = 0.110$
D	–	–	–	$P = 1.000$	$P = 1.000$
E	–	–	–	–	$P = 0.024$*
F	–	–	–	–	–

****P < 0.001; **P < 0.01; *P < 0.05.*

fatigue level 4 for six stimulus paradigms over 12 subjects. Two-way repeated-measures ANOVA revealed that the interaction of two factors of "stimulus paradigm" and "fatigue level" was non-significant in normalized α + θ index [$F_{(5,55)} = 1.930$, $P = 0.104$]. The factor of "stimulus paradigm" had a significant effect on α + θ index [Greenhouse–Geisser $F_{(2.466,27.128)} = 13.166$, $P < 0.001$], and the corresponding Bonferroni *post-hoc* analysis of α + θ index among six paradigms is shown in **Table 6**. The factor of "fatigue level" also had a significant effect on α + θ index [$F_{(1,11)} = 9.028$, $P = 0.012$]. The α + θ index and its change of paradigms A, B, and C were similar, revealing that paradigms A, B, and C had a similar anti-fatigue performance. Similarly, paradigms D and E also had a close anti-fatigue performance. Although the α + θ index of paradigm F was slightly higher than those of other paradigms at fatigue level 1, there was little change in the α + θ index between fatigue level 1 and level 4, demonstrating that paradigm F had better anti-fatigue efficacy than the other five paradigms. Hence, the anti-fatigue performance of six paradigms based on the α + θ index was as follows: F > A = B = C = E ≥ D.

DISCUSSION

In this study, we used four indices of α + θ index, pupil diameter index, NASA-TLX, and SSVEP amplitude and SNR to compare the mental fatigue and anti-fatigue performance of six paradigms in SSVEP visual acuity assessment. First, as for α + θ index, the θ waves tend to appear during meditative, drowsy, hypnotic, or sleeping states, and the increase in θ waves is related to performance decrements on task (Klimesch, 1999; Xie et al., 2016). The α waves appear during wakeful relaxation with closed eyes at decreased attention levels and in a drowsy but wakeful state, and the increase in α waves associated with fatigue is related to the increased mental effort to maintain vigilance level (Cao et al., 2014; Kathner et al., 2014). More specifically, the decreased attention and arousal level caused by mental fatigue are associated with the global increase in the θ and the α activities. Hence, the α + θ index shows a significant increase associated with the increasing fatigue level and also related to the mental alertness level (Eoh et al., 2005; Cao et al., 2014). Second, as for the pupil diameter, it is also a well-documented psycho-physiological proxy of effort, load on memory, and arousal (Peysakhovich et al., 2015), and the decrease in pupil diameter is related to deep breathing, mental work, and sleep. Hence, the increase of mental fatigue coincides with the decrease in pupil diameter (Hopstaken et al., 2015b). Third, NASA-TLX is a gold standard for measuring subjective mental fatigue across a wide range of applications (Hart and Staveland, 1988). Finally, previous studies have proved that the SSVEP amplitude and SNR can be significantly affected by the increasing fatigue level, and the amplitude and SNR of the elicited SSVEP are easily affected by mental states, fatigue, and degree of attention level (Cao et al., 2014; Xie et al., 2016).

This study focused on the mental fatigue effects caused by the long-time SSVEP stimulus of six stimulus paradigms. The results of all the indices of α + θ index, pupil diameter index, NASA-TLX, and SSVEP amplitude and SNR showed that paradigm F of motion expansion–contraction had a superior anti-fatigue efficacy than the other five paradigms of conventional onset mode or pattern reversal SSVEP stimulation during prolonged SSVEP experiment. The paradigm D of brief-onset mode showed the lowest anti-fatigue efficacy, and the other paradigms A, B, C, and E of pattern reversal SSVEP stimulation paradigms showed a similar anti-fatigue efficacy, which was between paradigms D and F. These indices of mental fatigue showed a good agreement. The results showed the anti-fatigue performance calculated averagely through all the four indices of mental fatigue estimation of six paradigms as follows: F > E ≥ A = B = C > D. Besides that, the decrease of SSVEP amplitude and SNR caused by mental fatigue during prolonged EEG experiment especially in paradigms A, B, C, D, and E may consequently deteriorate the SSVEP visual acuity assessment since the threshold determination criterion of SSVEP acuity is related to the amplitude and SNR of SSVEP response (Zheng et al., 2019a). Hence, we recommended paradigm F of oscillating expansion–contraction concentric rings as the stimulation paradigm in SSVEP visual acuity.

The reason for paradigm F to have the highest anti-fatigue property may be because of its uniform brightness and position changes rather than luminance alternations when presented to the subject, which overcame the problem of visual fatigue caused by uncomfortable light twinkling and contrast changes in the pattern reversal and brief-onset mode (Xie et al., 2012, 2016; Han et al., 2018; Yan et al., 2018). According to the theory of visual pathways, the visual system is divided into two major pathways of the parvocellular pathway and the magnocellular pathway (Pokorny and Smith, 1997). The magnocellular pathway contains the detection of dynamic motion and depth, whereas the parvocellular pathway contains the detection of spatial contrast and color information, with a slower propagation than the magnocellular pathway. Previous studies have proposed that attention uses the faster and more dominant signals of the magnocellular pathway to give priority to stimuli and simultaneously enhance the activity of the parvocellular pathway (Pokorny and Smith, 1997; Di Russo and Spinelli, 1999; Yeshurun and Sabo, 2012). If attentional networks are more reliant on parvocellular pathways, extra reaction time and demand are required for attention (Li et al., 2007; Laycock et al., 2008). Hence, the attention demand may be alleviated in motion expansion–contraction mode in paradigm F, while the contrast change in paradigm D of brief-onset mode may be a bit intense, resulting in the increase of attention demand.

There were also some limitations in this study that should be weighted. First, although the temporal frequency of 7.5 Hz was often used in SSVEP visual acuity (Almoqbel et al., 2011; Hamilton et al., 2020), the only one temporal frequency did not fully consider all relevant research since mental fatigue was limited by stimulus frequency to some extent (Won et al., 2016). Second, the number of trials and the time spent on one run were not necessarily the same as those of the SSVEP acuity test in clinical experiments, which may also have a certain influence on the results. Third, we used binocular rather than monocular viewing in our study, and the two may be not completely equivalent. Fourth, we used a consistent spatial frequency rather than a set of sweep spatial frequencies similar to the SSVEP visual acuity experiment, which may also lead to some difference in results from the actual experiment. Finally, in this study, compared to pupil diameter index and $\alpha + \theta$ index, there were no corresponding baseline of SSVEP amplitude and SNR in the first three pre-experimental trials of each paradigm because of no visual stimulus during the three trials, so SSVEP amplitude and SNR were not normalized by respective baseline.

CONCLUSION

To conclude, this study has explored the anti-fatigue performance of six stimulus paradigms (reverse vertical sinusoidal gratings, reverse horizontal sinusoidal gratings, reverse vertical square-wave gratings, brief-onset vertical sinusoidal gratings, reversal checkerboards, and oscillating expansion–contraction concentric rings) used in SSVEP visual acuity assessment. Four indices of $\alpha + \theta$ index, pupil diameter index, NASA-TLX, and SSVEP amplitude and SNR were proposed to estimate mental fatigue quantitatively. These indices of mental fatigue showed a good agreement. The results showed that the paradigm of motion expansion–contraction had a superior anti-fatigue efficacy than the other five paradigms of conventional onset mode or pattern reversal mode during prolonged SSVEP experiment. The paradigm of brief-onset mode showed the lowest anti-fatigue efficacy, and the other paradigms of pattern reversal mode showed a similar anti-fatigue efficacy, which was between motion expansion–contraction mode and onset mode.

Except for brief-onset vertical sinusoidal gratings, the four commonly used stimulus paradigms (i.e., reverse vertical sinusoidal gratings, reverse horizontal sinusoidal gratings, reverse vertical square-wave gratings, and reversal checkerboards) in SSVEP acuity assessment had a relatively good anti-fatigue property, indicating that mental fatigue could not affect the SSVEP acuity estimation too much when using the four stimulus paradigms. As for the paradigm of oscillating expansion–contraction concentric rings, it had the highest anti-fatigue property, and we recommended the oscillating expansion–contraction concentric rings as the stimulus paradigm in SSVEP acuity assessment.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the institutional review board of the Xi'an Jiaotong University. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

XZ contributed to the study design, data acquisition, analysis, interpretation, manuscript writing, and revision. GX contributed to the study design and the approval of the final version for publication. YZ contributed to the statistical analysis and manuscript drafting. RL and KZ contributed to the data analysis and interpretation. YD contributed to the manuscript writing and revision. JX provided the experimental equipment and approved the final version for publication. SZ conceptualized the study. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by grants from the Special Guidance Funds for the Construction of World-Class Universities (Disciplines) and the Characteristic Development in Central Universities (PY3A071), the National Key Research and Development Program of China (2017YFC1308500), and the National Natural Science Foundation of China (NSFC-51775415).

REFERENCES

Almoqbel, F. M., Yadav, N. K., Leat, S. J., Head, L. M., and Irving, E. L. (2011). Effects of sweep VEP parameters on visual acuity and contrast thresholds in children and adults. *Graefes. Arch. Clin. Exp. Ophthalmol.* 249, 613–623. doi: 10.1007/s00417-010-1469-8

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

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

Cao, T., Wan, F., Wong, C. M., da Cruz, J. N., and Hu, Y. (2014). Objective evaluation of fatigue by EEG spectral analysis in steady-state visual evoked potential-based brain-computer interfaces. *Biomed. Eng. Online* 13:28. doi: 10.1186/1475-925X-13-28

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

Chen, X., Wang, Y., Zhang, S., Xu, S., and Gao, X. (2019). Effects of stimulation frequency and stimulation waveform on steady-state visual evoked potentials using a computer monitor. *J. Neural Eng.* 16:066007. doi: 10.1088/1741-2552/ab2b7d

Di Russo, F., and Spinelli, D. (1999). Spatial attention has different effects on the magno- and parvocellular pathways. *Neuroreport* 10, 2755–2762. doi: 10.1097/00001756-199909090-00011

Eoh, H. J., Chung, M. K., and Kim, S. H. (2005). Electroencephalographic study of drowsiness in simulated driving with sleep deprivation. *Int. J. Ind. Ergonom.* 35, 307–320. doi: 10.1016/j.ergon.2004.09.006

Fahad, A., Leat, S. J., and Elizabeth, I. (2008). The technique, validity and clinical use of the sweep VEP. *Ophthal. Physl. Opt.* 28, 393–403. doi: 10.1111/j.1475-1313.2008.00591.x

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

Hamilton, R., Bach, M., Heinrich, S. P., Hoffmann, M. B., Odom, J. V., McCulloch, D. L., et al. (2020). VEP estimation of visual acuity: a systematic review. *Doc. Ophthalmol.* doi: 10.1007/s10633-020-09770-3

Han, C., Xu, G., Xie, J., Chen, C., and Zhang, S. (2018). Highly interactive brain-computer interface based on flicker-free steady-state motion visual evoked potential. *Sci. Rep.* 8:5835. doi: 10.1038/s41598-018-24008-8

Hart, S. G. (2006). Nasa-task load index (NASA-TLX); 20 years later. *Proc. Hum. Fact. Ergon. Soc. Annu. Meet.* 50, 904–908. doi: 10.1177/154193120605000909

Hart, S. G., and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. *Adv. Psychol.* 52, 139–183. doi: 10.1016/s0166-4115(08)62386-9

Hemptinne, C., Liu-Shuang, J., Yuksel, D., and Rossion, B. (2018). Rapid objective assessment of contrast sensitivity and visual acuity with sweep visual evoked potentials and an extended electrode array. *Invest. Ophthalmol. Vis. Sci.* 59, 1144–1157. doi: 10.1167/iovs.17-23248

Hopstaken, J. F., van der Linden, D., Bakker, A. B., and Kompier, M. A. (2015a). A multifaceted investigation of the link between mental fatigue and task disengagement. *Psychophysiology* 52, 305–315. doi: 10.1111/psyp.12339

Hopstaken, J. F., van der Linden, D., Bakker, A. B., and Kompier, M. A. (2015b). The window of my eyes: task disengagement and mental fatigue covary with pupil dynamics. *Biol. Psychol.* 110, 100–106. doi: 10.1016/j.biopsycho.2015.06.013

Kathner, I., Wriessnegger, S. C., Muller-Putz, G. R., Kubler, A., and Halder, S. (2014). Effects of mental workload and fatigue on the P300, alpha and theta band power during operation of an ERP (P300) brain-computer interface. *Biol. Psychol.* 102, 118–129. doi: 10.1016/j.biopsycho.2014.07.014

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res. Brain Res. Rev.* 29, 169–195. doi: 10.1016/s0165-0173(98)00056-3

Koo, B. Y., Jang, M. H., Kim, Y. C., and Mah, K. C. (2018). Changes in the subjective fatigue and pupil diameters induced by watching LED TVs. *Optik* 164, 701–710. doi: 10.1016/j.ijleo.2018.03.077

Laycock, R., Crewther, D. P., and Crewther, S. G. (2008). The advantage in being magnocellular: a few more remarks on attention and the magnocellular system. *Neurosci. Biobehav. Rev.* 32, 1409–1415. doi: 10.1016/j.neubiorev.2008.04.008

Lee, P. L., Sie, J. J., Liu, Y. J., Wu, C. H., Lee, M. H., Shu, C. H., et al. (2010). An SSVEP-actuated brain computer interface using phase-tagged flickering sequences: a cursor system. *Ann. Biomed. Eng.* 38, 2383–2397. doi: 10.1007/s10439-010-9964-y

Li, J. C., Sampson, G. P., and Vidyasagar, T. R. (2007). Interactions between luminance and colour channels in visual search and their relationship to parallel neural channels in vision. *Exp Brain Res.* 176, 510–518. doi: 10.1007/s00221-006-0773-3

Lin, Z., Zhang, C., Wu, W., and Gao, X. (2007). Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs. *IEEE Trans. Biomed. Eng.* 54, 1172–1176. doi: 10.1109/tbme.2006.889197

Listed, N. (2006). Guideline 5: guidelines for standard electrode position nomenclature. *Am. J. Eeg. Technol.* 46, 222–225. doi: 10.1080/1086508x.2006.11079580

Ng, K. B., Bradley, A. P., and Cunnington, R. (2012). Stimulus specificity of a steady-state visual-evoked potential-based brain-computer interface. *J. Neural Eng.* 9:036008. doi: 10.1088/1741-2560/9/3/036008

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

Norcia, A. M., and Tyler, C. W. (1985a). Infant VEP acuity measurements: analysis of individual differences and measurement error. *Electroencephalogr. Clin. Neurophysiol.* 61, 359–369. doi: 10.1016/0013-4694(85)91026-0

Norcia, A. M., and Tyler, C. W. (1985b). Spatial frequency sweep VEP: visual acuity during the first year of life. *Vis. Res.* 25, 1399–1408. doi: 10.1016/0042-6989(85)90217-2

Odom, J. V., Bach, M., Brigell, M., Holder, G. E., McCulloch, D. L., Mizota, A., et al. (2016). ISCEV standard for clinical visual evoked potentials: (2016 update). *Doc. Ophthalmol.* 133, 1–9. doi: 10.1007/s10633-016-9553-y

Peysakhovich, V., Dehais, F., and Causse, M. (2015). Pupil diameter as a measure of cognitive load during auditory-visual interference in a simple piloting task. *Proc. Manufact.* 3, 5199–5205. doi: 10.1016/j.promfg.2015.07.583

Pokorny, J., and Smith, V. C. (1997). Psychophysical signatures associated with magnocellular and parvocellular pathway contrast gain. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 14, 2477–2486. doi: 10.1364/josaa.14.002477

Sampei, K., Ogawa, M., Torres, C. C. C., Sato, M., and Miki, N. (2016). Mental fatigue monitoring using a wearable transparent eye detection system. *Micromachines* 7:20. doi: 10.3390/mi7020020

Tobimatsu, S., Kurita-Tashima, S., Nakayama-Hiromatsu, M., and Kato, M. (1993). Effect of spatial frequency on transient and steady-state VEPs: stimulation with checkerboard, square-wave grating and sinusoidal grating patterns. *J. Neurol Sci.* 118, 17–24. doi: 10.1016/0022-510x(93)90239-u

Tyler, C. W., Apkarian, P., Levi, D. M., and Nakayama, K. (1979). Rapid assessment of visual function: an electronic sweep technique for the pattern visual evoked potential. *Invest. Ophthalmol. Vis. Sci.* 18, 703–713.

Won, D. O., Hwang, H. J., Dahne, S., Muller, K. R., and Lee, S. W. (2016). Effect of higher frequency on the classification of steady-state visual evoked potentials. *J. Neural Eng.* 13:016014. doi: 10.1088/1741-2560/13/1/016014

Wu, Z., Yao, D., Tang, Y., Huang, Y., and Su, S. (2010). Amplitude modulation of steady-state visual evoked potentials by event-related potentials in a working memory task. *J. Biol. Phys.* 36, 261–271. doi: 10.1007/s10867-009-9181-9

Xie, J., Xu, G., Luo, A., Li, M., Zhang, S., Han, C., et al. (2017). The role of visual noise in influencing mental load and fatigue in a steady-state motion visual evoked potential-based brain-computer interface. *Sensors* 17:1873. doi: 10.3390/s17081873

Xie, J., Xu, G., Wang, J., Li, M., Han, C., and Jia, Y. (2016). Effects of mental load and fatigue on steady-state evoked potential based brain computer interface tasks: a comparison of periodic flickering and motion-reversal based visual attention. *PLoS One* 11:e0163426. doi: 10.1371/journal.pone.0163426

Xie, J., Xu, G., Wang, J., Zhang, F., and Zhang, Y. (2012). Steady-state motion visual evoked potentials produced by oscillating Newton's rings: implications for brain-computer interfaces. *PLoS One* 7:e39707. doi: 10.1371/journal.pone.0039707

Yadav, N. K., Almoqbel, F., Head, L., Irving, E. L., and Leat, S. J. (2009). Threshold determination in sweep VEP and the effects of criterion. *Doc. Ophthalmol.* 119, 109–121. doi: 10.1007/s10633-009-9177-6

Yan, W., Xu, G., Chen, L., and Zheng, X. (2019). Steady-state motion visual evoked potential (SSMVEP) enhancement method based on time-frequency image fusion. *Comput. Intell. Neurosci.* 2019:9439407. doi: 10.1155/2019/9439407

Yan, W., Xu, G., Xie, J., Li, M., and Dan, Z. (2018). Four novel motion paradigms based on steady-state motion visual evoked potential. *IEEE Trans. Biomed. Eng.* 65, 1696–1704. doi: 10.1109/TBME.2017.2762690

Yeshurun, Y., and Sabo, G. (2012). Differential effects of transient attention on inferred parvocellular and magnocellular processing. *Vis. Res.* 74, 21–29. doi: 10.1016/j.visres.2012.06.006

Zheng, X., Xu, G., Wang, Y., Han, C., Du, C., Yan, W., et al. (2019a). Objective and quantitative assessment of visual acuity and contrast sensitivity based on steady-state motion visual evoked potentials using concentric-ring paradigm. *Doc. Ophthalmol.* 139, 123–136. doi: 10.1007/s10633-019-09702-w

Zheng, X., Xu, G., Zhi, Y., Wang, Y., Han, C., Wang, B., et al. (2019b). Objective and quantitative assessment of interocular suppression in strabismic amblyopia based on steady-state motion visual evoked potentials. *Vis. Res.* 164, 44–52. doi: 10.1016/j.visres.2019.07.003

Zheng, X., Xu, G., Wu, Y., Wang, Y., Du, C., Wu, Y., et al. (2020). Comparison of the performance of six stimulus paradigms in visual acuity assessment based on steady-state visual evoked potentials. *Doc. Ophthalmol.* doi: 10.1007/s10633-020-09768-x

Zhu, D., Bieger, J., Garcia Molina, G., and Aarts, R. M. (2010). A survey of stimulation methods used in SSVEP-based BCIs. *Comput. Intell. Neurosci.* 2010:702357. doi: 10.1155/2010/702357

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

frontiers in Human Neuroscience

METHODS
published: 18 September 2020
doi: 10.3389/fnhum.2020.00365

Bi-Dimensional Approach Based on Transfer Learning for Alcoholism Pre-disposition Classification via EEG Signals

Hongyi Zhang[1], Francisco H. S. Silva[2], Elene F. Ohata[2,3], Aldisio G. Medeiros[2,3] and Pedro P. Rebouças Filho[2,3,4]*

[1] School of Opto-Electronic and Communication Engineering, Xiamen University of Technology, Xiamen, China, [2] Laboratório de Processamento de Imagens, Sinais e Computação Aplicada, Instituto Federal do Ceará, Fortaleza, Brazil, [3] Programa de Pós-Graduação em Engenharia de Teleinformática, Universidade Federal do Ceará, Fortaleza, Brazil, [4] Programa de Pós-Graduação em Ciência da Computação, Instituto Federal do Ceará, Fortaleza, Brazil

Recent statistics have shown that the main difficulty in detecting alcoholism is the unreliability of the information presented by patients with alcoholism; this factor confusing the early diagnosis and it can reduce the effectiveness of treatment. However, electroencephalogram (EEG) exams can provide more reliable data for analysis of this behavior. This paper proposes a new approach for the automatic diagnosis of patients with alcoholism and introduces an analysis of the EEG signals from a two-dimensional perspective according to changes in the neural activity, highlighting the influence of high and low-frequency signals. This approach uses a two-dimensional feature extraction method, as well as the application of recent Computer Vision (CV) techniques, such as Transfer Learning with Convolutional Neural Networks (CNN). The methodology to evaluate our proposal used 21 combinations of the traditional classification methods and 84 combinations of recent CNN architectures used as feature extractors combined with the following classical classifiers: Gaussian Naive Bayes, K-Nearest Neighbor (k-NN), Multilayer Perceptron (MLP), Random Forest (RF) and Support Vector Machine (SVM). CNN MobileNet combined with SVM achieved the best results in Accuracy (95.33%), Precision (95.68%), F1-Score (95.24%), and Recall (95.00%). This combination outperformed the traditional methods by up to 8%. Thus, this approach is applicable as a classification stage for computer-aided diagnoses, useful for the triage of patients, and clinical support for the early diagnosis of this disease.

Keywords: electroencephalogram, alcoholism, convolutional neural network, computer vision, transfer learning

OPEN ACCESS

Edited by:
Adel Said Elmaghraby, University of Louisville, United States

Reviewed by:
Chee-Ming Ting, University of Technology Malaysia, Malaysia
Fuad Noman, Universiti Tenaga Nasional, Malaysia

****Correspondence:***
Hongyi Zhang
zhanghongyi@xmut.edu.cn

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

Received: *13 May 2020*
Accepted: *10 August 2020*
Published: *18 September 2020*

Citation:
Zhang H, Silva FHS, Ohata EF, Medeiros AG and Rebouças Filho PP (2020) Bi-Dimensional Approach Based on Transfer Learning for Alcoholism Pre-disposition Classification via EEG Signals. Front. Hum. Neurosci. 14:365. doi: 10.3389/fnhum.2020.00365

1. INTRODUCTION

In 2016, there were around 3 million deaths worldwide due to alcohol abuse, 5.3% of all deaths recorded that year. The number of deaths from alcohol is greater than from some other serious diseases like tuberculosis, Acquired Immunodeficiency Syndrome (AIDS) and diabetes (World Health Organization, 2019). Still in 2016, alcohol caused 132.6 million disability-adjusted life years (DALYs) which represented 5.1% of all DALYs in that year. The World Health Organization (WHO) estimates that 283 million people worldwide have alcohol use disorders (World Health Organization, 2019).

Moderate and frequent alcohol consumption (>30 g/day) can bring benefits to the cardiovascular system (Foppa et al., 2001), with increased High Density Lipoprotein (HDL) cholesterol and the consumption of red wine has antioxidant action (da Luz and Coimbra, 2004). However, abusive alcohol consumption (>60 g/day) has direct consequences for the medium-long term health of the individual, such as liver disease, cancer, cardiovascular, and mental problems, as well as indirect consequences in case of accidents, suicides, and homicides due to short-term harm, such as cognitive and mobility problems (da Luz and Coimbra, 2004; Jennison, 2004; World Health Organization, 2019). Alcohol affects the Central Nervous System (CNS) directly, causing changes in its function and in brain functions. One way to check brain activity and the changes caused by alcohol is through an EEG exam (Devor and Cloninger, 1989) which can identify different types of brain activities through electrodes placed on specific regions of the head.

The EEG has multiple channels to collect electrical signals, which are emitted through neuron synapses and are incredibly complicated and non-linear. Specific techniques are required to interpret the complexity of these exams. Computer Aided Diagnostic (CAD) tools along with the use of Digital Signal Processing (DSP) and Artificial Intelligence (AI) techniques, with Machine Learning (ML) (Acharya et al., 2012; McBride et al., 2015; Patidar et al., 2017; Bhattacharyya et al., 2018; Bosl et al., 2018; Ibrahim et al., 2018; Amezquita-Sanchez et al., 2019; Rodrigues et al., 2019) can be applied to interpret these signals. Several studies have been performed using EEG signals to identify different types of disturbances in brain activity, such as the detection of patterns that characterize Alzheimer's disease (McBride et al., 2015; Amezquita-Sanchez et al., 2019; Tzimourta et al., 2019), autism (Boutros et al., 2015; Bosl et al., 2018; Ibrahim et al., 2018), sleep disorders (Koley and Dey, 2012; D'Rozario et al., 2015; Rundo et al., 2019), hyperactivity (Mohammadi et al., 2016; Muñoz-Organero et al., 2018; Wang et al., 2019), and epilepsy (Bhattacharyya et al., 2018; Ibrahim et al., 2018; Ren et al., 2018).

Besides the fact that EEG exams have previously presented good results in identifying different diseases, we chose to use the EEG exam because it provides an extensive mapping of brain activity equal to other exams, such as Magnetoencephalography (MEG), functional Magnetic Resonance Imaging (fMRI), functional Near-Infrared Spectroscopy (fNIRS) and Positron Emission Tomography (PET). However, recording EEG signals is simpler than MEG signals, since the measurement of electrical voltages is more easily performed than the measurement of magnetic fields as they have a low amplitude (Stam, 2010). Hair artifacts can influence infrared-based fNIRS measurements, and thus directly interfering with the reliability of the exam (Lloyd-Fox et al., 2010). EEG does not emit particles to obtain the result of the examination, as in the case of PET (Chugani et al., 1987). Furthermore, fMRI (Kozel et al., 2004) requires the use of high-cost magnetic scanners unlike EEG, which in comparison is a low-cost equivalent solution.

1.1. Contribution and Paper Organization

Among the main contributions of this work to diagnose a predisposition to alcoholism, we highlight the use of a heat map to represent the brain activity of each patient in order to provide a visual analysis and the use of the Transfer Learning method, as the extraction of deep attributes as a way to represent the healthy and pathologic samples.

The paper is organized as follows: section 2 presents a literature review concerning the topic. Section 3 discusses the materials and methods that support the proposed technique. Section 4 gives a description of the use of CNN as an attribute extractor. The proposed methodology is described in section 5, and finally, in section 6, we present the results obtained and the discussion.

2. OVERVIEW OF THE ALCOHOLISM PREDISPOSITION CLASSIFICATION

This section presents the state of the art of EEG analysis to identify alcoholism considering the evolution of feature extraction methods from the traditional statistical approach to the current use of CNNs as feature extractors.

Acharya et al. (2012) developed an automatic technique for CAD to identify healthy patients with a genetic predisposition to alcoholism through EEG signal analyses. These authors combined a non-linear feature extraction, such as Approximate Entropy, Sample Entropy, Largest Lyapunov Exponent, and four Higher-Order Spectra (HOS) functions with a SVM classifier, varied the Polynomial and Radial Basis Function (RBF) kernals. Their results indicated that non-linear measurements extracted from EEG signals can achieve promising results.

Using the electrical impulses that represent the physiological functions like eye blinking and heart beating, Rachman et al. (2016) proposed an independent component analysis through EEG signals. In their work, the features extracted by stationary wavelet transform with Daubechies decomposition at level 6 were combined with a probabilistic neural network to classify samples from 64 channels into two classes: healthy and alcoholism patients. However, this work only used classical statistic features like maximum, minimum, and average values, showing its fragility when outlier samples were present in the dataset.

Mumtaz et al. (2016), on the other hand, analyzed 19 channels placed according to the international 10–20 system to identify healthy and alcoholism patient. The dataset had 18 alcoholism and 15 healthy patients. They extracted features through quantitative electroencephalography from EEG data. The features were used as the input for classification models: Linear Discriminant Analysis, SVM, MLP, and Logistic Model Trees. This study suggests that EEG spectral analysis can help to classify pathologic samples from the healthy ones. Nevertheless, they used seven frequency bands in these analyses, indicating an increase in the time to generate results.

Ehlers et al. (1998) proposed an approach to evaluate the influence of alcohol consumed on brain activities. They analyzed EEG signals through temporal series combined with the chaos theory. In their study, the authors assessed two groups of patients,

a control group, and an alcoholism group. Based on this, they suggested that the EEG signal has non-linear structures that can be modified when the patient is under the effect of alcohol.

Kannathal et al. (2005) analyzed EEG signals through non-linear measurements, such as correlation dimension, largest Lyapunov exponent, Hurst exponent, and entropy values. The authors suggested that non-linear analysis could contribute to distinguish between healthy and alcoholic patients.

Faust et al. (2013) also considered the non-linear characteristics of EEG signals. These authors used the non-linear feature of HOS to extract information about alcoholic patients. This feature was used as the input to six different classifications models: Decision Tree, Fuzzy Sugeno Classifier, k-NN, Gaussian Mixture Model, Naive Bayes Classifier, and Probabilistic Neural Network.

Although these recent works in the literature have presented promising results, some of them omitted the number of samples evaluated and which criteria was used to select the channels of the EEG exam to be analyzed. Furthermore, most of these works use feature extraction techniques specially adjusted to assessed datasets, hindering the possibility to generalize to signals with other characteristics. Finally, these works did not evaluate new feature extractors, especially such algorithms based on the recent technique of Deep Learning (DL) using Transfer Learning; this is one of the innovations of our approach.

Moreover, these works were performed using the raw one-dimensional signals of the EEG, in addition to selecting specific channels to solve the problem. In our work, we proposed a two-dimensional heat-map representations to represent the EEG channels, where each value acquired from one channel corresponds to the pixel value in the resulting image, so the junction of all selected channels makes up the final image for each patient.

The generated image corresponds to the heat map of the brain activity of this patient, thus giving a visual analysis of the problem, as well as the use of CV, DL, and ML methods. The use of heat map imaging enables the application of structural and textural analysis methods, such as pixel variance, morphological gradient calculations, equalization, as well as enhancement algorithms that can improve the distinction between alcoholic and healthy samples; thus giving a more accurate diagnostic.

The two-dimensional approach also allows the use of feature extraction methods, which describe different shapes, textures and structures of each image, such as Gray-Level Co-Occurrence Matrix (GLCM) (Haralick et al., 1973), Hu's Moments (Hu, 1962), and Local Binary Patterns (LBP) (Ojala et al., 2002). Furthermore, the application of the Transfer Learning technique using CNNs enables the extraction of the most relevant features from an image through extreme non-linear models. The classification of these characteristics belonging to each patient is obtained using ML algorithms. Through a Random Search for the optimal parameters, we obtained the best configuration of the following models: k-NN (Fukunaga and Narendra, 1975), MLP (Haykin, 2008), RF (Breiman, 2001), and SVM (Vapnik, 1998).

3. MATERIALS

In this section, we present the digital image processing techniques and the ML that supports the methodology proposed in this work.

3.1. Dataset

The dataset used in this work is publicly available in Begleiter (2019) from the University of California, Irvine, and is known as Knowledge Discovery in Database (UCI KDD). This dataset was initially developed to examine genetic predisposition, through EEG signals, to alcoholism. Two subject groups made up the dataset: an alcoholics group and a control group. The Alcoholic group consists of the 77 male subjects with a mean age of 35.83 $\pm$ 5.33. The control group consists of 48 male subjects with no particular or family history of alcohol misuse or neurological disorder or any history of psychiatric disease.

The signal acquisition is according to the 10–20 International System with 64 electrodes placed on the scalps of the subjects, with a sampling frequency of 256 samples per second. The Cz electrode is taken as a reference. Each signal has a period of 190 ms of a pre-stimulation and 1,440 ms after each stimulus.

Each subject was exposed to three conditions, a single stimulus (S1) was presented to each subject. A second stimulus (S2) is a matching condition, here the same stimulus S1 was repeated. Finally, the last stimulus (S3) presented in either a matched condition where S1 was identical to S2. Each stimulus corresponds to a picture of objects chosen from the 1980 Snodgrass and Vanderwart picture set (Snodgrass and Vanderwart, 1980).

3.2. Tradictional Feature Extraction Methods

In this study, three feature extraction methods were used to improve the analysis of the proposed approach.

Haralick et al. (1973) proposed a statistical analysis considering the co-occurrence of gray levels in the image. This method is called Gray-Level Co-Occurrence Matrix (GLCM) and identifies the spatial influences of pixels related to their grayscale. GLCM has 14 features, and among which the angular second moment and entropy are commonly used and here they are presented in Equations (1) and (2), respectively, where p is central pixel, i and j are indexes according to image height and width.

$$\sum_i \sum_j p(i,j)^2 \tag{1}$$

$$-\sum_i \sum_j p(i,j) log(p(i,j)) \tag{2}$$

The Local Binary Patterns (LBP) proposed by Ojala et al. (1994), was developed as an efficient and straightforward way to describe the texture of an image. LBP extracts information from the local gray scale levels of the image to define a pattern that represents

P pixels of the near neighbors. This binary pattern follows a pattern determined by neighbors analysis direction. Equation (3) presents the neighborhood analysis, where g_p is a neighbor pixel P to the region of the radius R, and g_c is the central pixel.

$$LBP_{P,R} = \sum_{p=0}^{P-1} f(g_p - g_c)^{2P} \tag{3}$$

According to the threshold x, a binary pattern is assigned to each operation (Equation 4).

$$f(x) = \begin{cases} 1, & if \quad x \geq 0 \\ 0 & otherwise \end{cases} \tag{4}$$

Hu (1962) developed a model that uses central moments to make the method invariant to scale and rotational changes. This method, known as HU moments, describes a feature extraction family composed of seven moments; each one is invariable to size, rotation, and translation operations. Equation (5) shows the relation between central moment and normalized moment. This normalized moment can be obtained from the central moment, μ_{pq}, divided by an exponential of the area, μ_{00}, to obtain the normalized central moment, η_{pq}.

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\alpha}} \tag{5}$$

where

$$\alpha = \frac{p+q}{2}; \forall p+q \geq 2 \tag{6}$$

3.3. Classifiers

This section describes the ML techniques used to classify the features extracted by the traditional methods and the CNN architectures.

3.3.1. Naive Bayes

Bayesian Classifier is based on statistical analysis of input data. The classifications are based on the probability distribution of each sample to a specific class, considering that this class has the highest probability to be associated with the sample (Theodoridis and Koutroumbas, 2008). The Bayes Theory inspires this model, and it assumes that there are no dependencies among the features, according to the value of posterior probability and conditional probability.

3.3.2. K-Nearest Neighbor

K-Nearest Neighbor (k-NN) is a machine learning method proposed by Fukunaga and Narendra (1975) that falls into the supervised category. It determines the class to which a sample belongs by comparing the features of the k nearest neighbors that were acquired in a previous training step. The variable k represents the number of samples of the training set that possess the closest features to the sample being classified. Still regarding the variable k, there is not a standard value for it, but in general, even numbers are avoided to prevent a drawn situation in which the sample could be classified into two classes at the same time.

3.3.3. Multilayer Perceptron—MLP

Multilayer Perceptron (MLP) is a neural network architecture formed by multiple layers with perceptron neurons. The input data vector is introduced to the first layer where each feature is computed and each neuron contributes to transform the input space into a linearly separable space and thus to classify the object in its specific class (Haykin, 2008). The learning technique is supervised through a backpropagation algorithm where the errors calculated at the last layer are retro propagated to adjust the hidden layers (Haykin, 2008). Therefore, throughout this procedure, the solution to samples in the input vector is presented in the output layer.

3.3.4. Random Forest

Random Forest (RF) is based on decision trees, proposed by Breiman (2001). It aims to make a decision tree using a set of features selected from the initial set. The training is achieved by using a meta-algorithm called bagging, which uses the stability and accuracy of the results to improve the classification. Bagging is used to reduce the variance and over-fitting. After the tree sets are created, it is possible to determine which set contains the best configuration to solve a problem.

3.3.5. Support Vector Machine—SVM

A Support Vector Machine (SVM) is based on the statistic distribution of the samples in the input vector proposed by Suykens and Vandewalle (1999). SVMs aim to identify samples that are most difficult to classify because they are close to the decision boundary. This method uses the optimization theory to adjust the optimal decision boundary for the minimization of the cost function with restriction parameters. Originally developed for binary classification, this classifier can be extended to multiclass problems through the one-against-all and one-against-one approaches, in addition these are techniques based on the graph theory (Vapnik, 1998). SVMs can be applied to both linear and non-linear problems, this latter method can use an RBF type kernel.

4. CONVOLUTIONAL NEURAL NETWORKS

In this paper, we evaluated the following CNNs: DenseNet (Huang et al., 2017), Inception-ResNet (Längkvist et al., 2014), Inception (Szegedy et al., 2015), MobileNet (Howard et al., 2017), NasNet (Zoph and Le, 2016), ResNet (Wu et al., 2018), VGG (Simonyan and Zisserman, 2014), and Xception (Chollet, 2017).

4.1. Convolutional Neural Networks as Feature Extractor

In this paper, CNNs used the transfer learning concept, which relates the descriptive power of a pre-trained CNN on samples of a problem not yet known by the model. The first fully connected layer is removed, and then, a resizing of its input is transformed into a one-dimensional array. After this process, a pre-trained

model does not behave as a classifier, so it is used as a feature extractor. The transfer learning technique is detailed in the work of da Nóbrega et al. (2018), who applied transfer learning to lung nodule classification.

4.1.1. Architecture Construction and Initialization

Many architectures have been proposed in the last few years, especially since 2010, with the advent of object recognition challenges in large scale image datasets (Deng et al., 2009). However, it is not viable to evaluate all of the architectures proposed by the scientific community; therefore, 12 well-known architectures were selected for the experiments of this work. The configurations of the models described in their respective paper were used during implementation.

4.1.2. Architecture Training

The twelve architectures were trained from the ImageNet dataset (Deng et al., 2009), which consists of 1.2 million non-medical images, and grouped into 1,000 categories. The training methodologies used by each architecture are documented in detail in their respective articles. This step was done based on the premise that the features learned by a CNN to discriminate a set

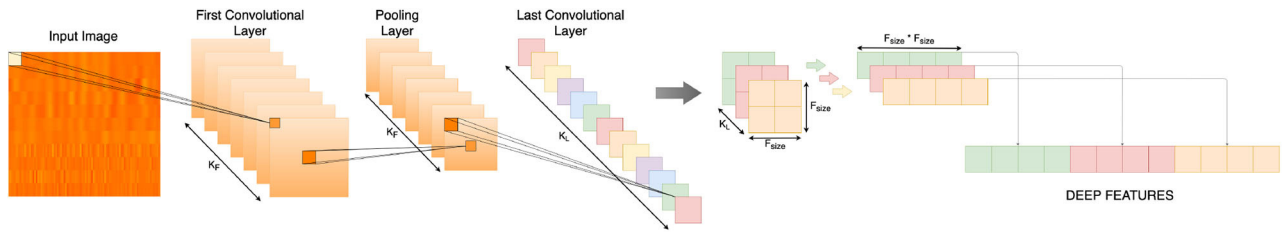

FIGURE 1 | Transfer learning figure.

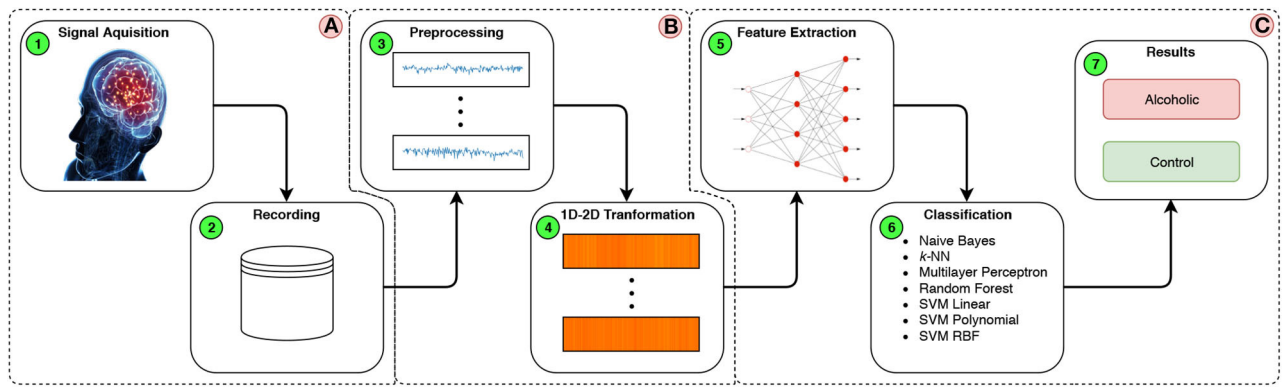

FIGURE 2 | Flow chart of the proposed methodology.

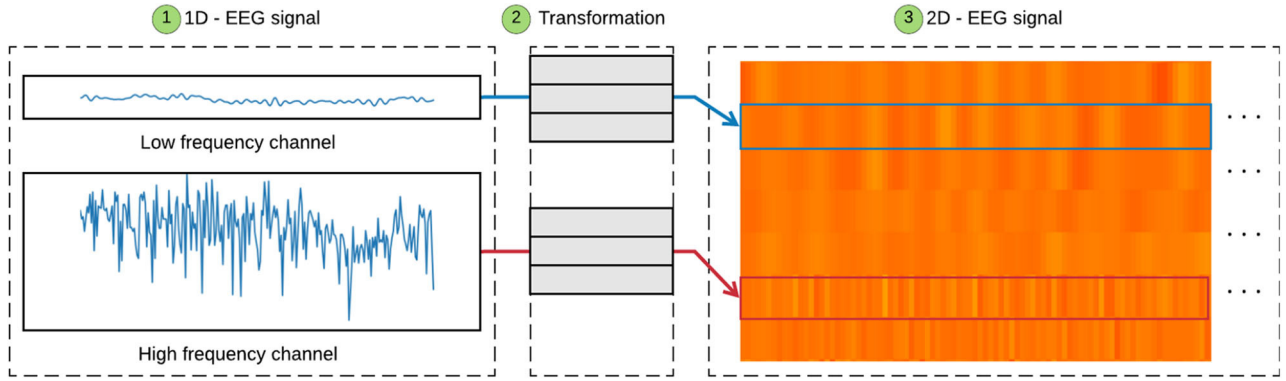

FIGURE 3 | The transformation from 1D EEG signals to 2D EEG signals. Low-frequency channels are transformed to a smooth texture block. High-frequency channels are transformed to a rough texture block.

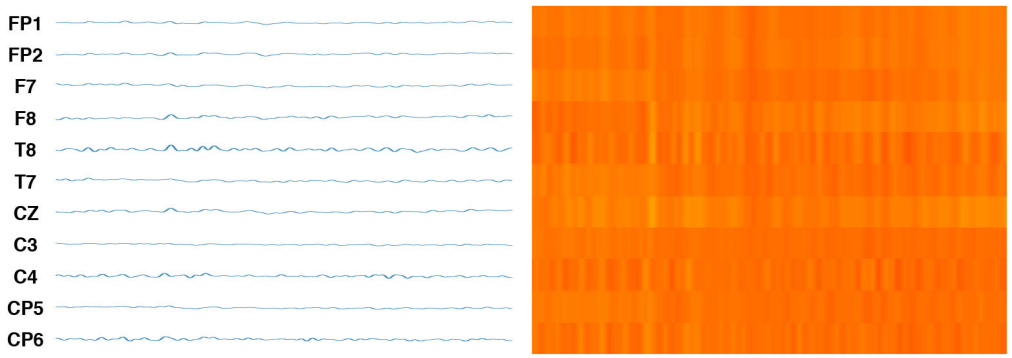

FIGURE 4 | On the left, original EEG signals. On the right the final image after channels transformation from the 1D channels to a 2D image.

of classes are capable of representing other samples as the model can extrapolate the known patterns to new sample with the use of transfer learning technique.

4.1.3. Converting CNNs Into Feature Extractor

In this last step, the CNNs trained on the previously mentioned set are transformed into feature extractors. Nonetheless, to perform this step, it is crucial to understand the four transformations executed by these neural networks.

Initially, the input image is submitted to a sequence of non-linear transformations. These transformations are defined depending on the architecture used. In this first stage, the input image is converted into a set of small matrices. Secondly, each of these matrices is resized to a one-dimensional array. Then, the set of arrays is concatenated, thus generating a single array. Each one-dimensional array can be interpreted as a feature vector that represents the heat map image. Lastly, the features vectors are submitted to a classifier training. With the modified architecture, the results of the model should not be interpreted as a probability set of an input image related to a determined label but should be interpreted as an information vector, which will be used by an external classifier to compose the probabilities of predisposition to alcoholism.

Figure 1 shows the fully connected layer after the removal of the last convolutional layer. The outputs are concatenated and then the vectors set that will be used to train and test the classifier are created.

TABLE 1 | Number of features returned by each extractor.

Approach	Extractor	Number of features
Traditional	LBP	48
	GLCM	14
	Hu Moments	7
Transfer Learning	Densenet 121	1,024
	Densenet 169	1,664
	Densenet 201	1,920
	InceptionResNetV2	1,536
	InceptionV3	2,048
	MobileNet	1,024
	NASNetLarge	4,032
	NASNetMobile	1,056
	ResNet50	2,048
	VGG16	512
	VGG19	512
	Xception	2,048

5. METHODOLOGY

In this paper, we propose the detection of a predisposition to alcoholism comparing EEG signals from two subgroups: alcoholism and control. **Figure 2** illustrates the proposed methodology step-by-step and it is divided into three main stages: Acquisition of the EEG signals (A), Digital Signal Processing (DSP) (B), and finally, extraction and classification of the samples (C).

5.1. Pre-processing Step

Out of the 64 exam channels, only 11 were selected from the mean-variance of each channel for all patients in the dataset. The selected channels were: FP1, FP2, F7, F8, T8, T7, T7, CZ, C3, C4, CP5, and CP6. These channels presented the highest values of variance in their signals, which means more intense brain activity in the regions where these channels were located.

Initially, stage A was performed out during the formation of the Dataset. In stage B the data is prepared in step 3 (**Figure 2**B-3) by removing any outliers, >73.3 and < -73.3 uV, which represent possible head and eye movements (Zhang et al., 1995), and then the set of signals is normalized within a range of 0–1. In step 4 (**Figure 2**B-4), the interval is readjusted to 0–255, in addition to turning all values into integers, which enables the creation of an 8-bit image with $1{,}024 \times 352$ shape that represents the concatenation of the exam channels, where each of the selected channels are $1{,}024 \times 32$ pixel regions, as shown in **Figure 3**.

Finally, in stage C, the CNN technique is used as a feature extractor combined with a Transfer Learning method. The extracted features are classified in Alcoholic and Control, using traditional classification models.

Figure 3 shows the transformation process from a 1D EEG signal to a 2D EEG signal, highlighting the distinction between low and high-frequency. Step 1 (**Figure 3**-1) shows the 1D EEG signals. The channels are transformed into a 2D image in step 2 (**Figure 3**-2) as previously described. Step 3 (**Figure 3**-3) shows that high-frequency signals are represented by roughly textured blocks, creating a surface with peaks and valleys due to the high variation of the signals. On the other hand, low-frequency signals are represented by smooth texture blocks, presenting a flat surface due to the low variation of the signals. We found such signals, high and low frequency, using the calculation of the mean-variance of the exam channels.

Finally, **Figure 4** shows a sample of the complete transformation from the 1D channels to a 2D image. This approach becomes a visual representation of brain activity in different parts of the brain, rather than treating each channel separately. This image corresponds to a heat map of the brain activity in the regions measured by the electrodes. This image clearly represents the variations of time-domain and reflect the temporal variations of the channels through image texture, as well as the EEG signal by the intensity of color. This approach permits a visual analysis of the problem, as well as the use of structural and textural analytical methods. Moreover, the approach makes it possible to use recent methods of CV, DL, and ML methods.

5.2. Feature Extraction

In **Figure 2** stage 2-C, step 5 (**Figure 2**C-5), according to section 4.1.3, this approach proposes the use of CNN techniques as a feature extractor combined with a Transfer Learning method for the two-dimensional signals. The image that represents the EEG signals is processed by the first convolutional layers of the neural network, and the output of the final layer of the CNN is combined as a feature vector for the classification stage. The traditional computational vision extractors, such as GLCM, Hu Moments, and LBP were evaluated. **Table 1** shows the number of features generated by each extractor.

5.3. Classification Healthy and Alcoholism Patient

To evaluate the representativeness of the extracted features for the classification of both sets, healthy patients and alcoholic patients, the generated dataset is classified using five consolidated ML techniques: Bayes (Theodoridis and Koutroumbas, 2008), k-NN (Fukunaga and Narendra, 1975), RF (Breiman, 2001), MLP (Haykin, 2008), and SVM (Vapnik, 1998).

In the classification process, Bayes classifier operated with the Probability Density Function (PDF). MLP performed its training using the Levenberg-Marquardt method, and with the neurons varying from 2 to 1,000 in the hidden layer. The number of neighbors for the k-NN classifier was determined through a grid search, where the k value was varied using the odd values from 3 to 15.

The SVM classifier used linear, polynomial and RBF kernels. In all three configurations, the C hyperparameter was defined as 2^{-5}, 2^{-4}, 2^{-3}, …, 2^{15}. For the RBF kernel, γ was varied from 2^{-15} to 2^{3}, while for polynomial kernel, the degree ranged using the odd values from 3 to 9.

TABLE 2 | Accuracy, Precision, F1-Score, and Recall obtained through the classification of extracted features with classical extractors.

Extractors	Classifiers	Accuracy	Precision	F1 Score	Recall
GLCM	Naive Bayes	64.44 ± 1.22	81.61 ± 0.40	46.55 ± 2.96	54.29 ± 1.56
	MLP	64.44 ± 1.22	81.61 ± 0.40	46.55 ± 2.96	54.29 ± 1.56
	kNN	86.11 ± 3.23	87.70 ± 3.97	84.66 ± 3.69	83.55 ± 3.73
	RF	**87.22 ± 2.87**	**89.06 ± 3.40**	**85.86 ± 3.18**	**84.56 ± 3.12**
	SVM Linear	73.22 ± 3.12	78.87 ± 4.39	66.47 ± 4.96	66.71 ± 4.05
	SVM Polynomial	64.44 ± 1.22	81.61 ± 0.40	46.55 ± 2.96	54.29 ± 1.56
	SVM RBF	72.22 ± 2.77	84.11 ± 1.14	62.83 ± 4.91	64.34 ± 3.62
HU	Naive Bayes	61.11 ± 0.00	30.56 ± 0.00	37.93 ± 0.00	50.00 ± 0.00
	MLP	61.11 ± 0.00	30.56 ± 0.00	37.93 ± 0.00	50.00 ± 0.00
	kNN	80.44 ± 3.19	81.64 ± 4.05	78.14 ± 3.85	77.19 ± 3.74
	RF	**80.67 ± 3.30**	**81.59 ± 4.08**	**78.51 ± 3.82**	**77.58 ± 3.78**
	SVM Linear	52.89 ± 3.79	55.34 ± 3.52	52.81 ± 3.78	55.38 ± 3.64
	SVM Polynomial	51.89 ± 2.77	55.72 ± 2.91	51.75 ± 2.87	55.44 ± 2.84
	SVM RBF	50.56 ± 4.28	57.97 ± 3.80	49.31 ± 5.34	56.17 ± 3.54
LBP	Naive Bayes	61.11 ± 0.00	30.56 ± 0.00	37.93 ± 0.00	50.00 ± 0.00
	MLP	61.11 ± 0.00	30.56 ± 0.00	37.93 ± 0.00	50.00 ± 0.00
	kNN	83.89 ± 2.87	84.81 ± 4.13	82.41 ± 2.89	81.47 ± 2.64
	RF	**87.33 ± 3.82**	**89.08 ± 3.99**	**85.96 ± 4.35**	**84.75 ± 4.49**
	SVM Linear	66.89 ± 3.33	65.00 ± 3.94	63.21 ± 3.81	63.09 ± 3.55
	SVM Polynomial	38.89 ± 0.00	19.44 ± 0.00	28.00 ± 0.00	50.00 ± 0.00
	SVM RBF	68.56 ± 2.33	71.01 ± 5.78	60.57 ± 3.66	61.70 ± 2.77

The bold values are mean and standard deviation, respectively.

TABLE 3 | Accuracy, Precision, F1-Score, and Recall obtained through the classification of extracted features with CNNs architectures.

Extractors	Classifiers	Accuracy	Precision	F1 Score	Recall
DenseNet121	Naive Bayes	72.78 ± 1.88	80.75 ± 2.42	64.94 ± 3.36	65.57 ± 2.48
	MLP	75.78 ± 2.37	79.59 ± 4.55	71.66 ± 4.66	71.51 ± 3.88
	kNN	87.44 ± 2.44	89.13 ± 2.56	86.12 ± 2.86	84.90 ± 2.93
	RF	85.22 ± 2.81	87.35 ± 3.09	83.44 ± 3.35	82.09 ± 3.35
	SVM Linear	89.67 ± 3.65	90.21 ± 4.20	88.89 ± 3.90	88.17 ± 3.85
	SVM Polynomial	38.89 ± 0.00	19.44 ± 0.00	28.00 ± 0.00	50.00 ± 0.00
	SVM RBF	**89.78 ± 3.51**	**90.32 ± 4.05**	**89.01 ± 3.74**	**88.31 ± 3.69**
DenseNet169	Naive Bayes	71.00 ± 4.26	77.11 ± 7.90	62.22 ± 6.80	63.55 ± 5.00
	MLP	72.56 ± 6.05	77.58 ± 8.26	65.86 ± 8.90	66.64 ± 7.39
	kNN	87.22 ± 2.78	88.30 ± 3.10	86.04 ± 3.22	85.13 ± 3.38
	RF	86.56 ± 3.67	89.05 ± 3.23	84.85 ± 4.48	83.44 ± 4.51
	SVM Linear	89.11 ± 5.01	90.23 ± 5.05	88.05 ± 5.63	87.09 ± 5.84
	SVM Polynomial	38.89 ± 0.00	19.44 ± 0.00	28.00 ± 0.00	50.00 ± 0.00
	SVM RBF	**91.33 ± 3.36**	**92.26 ± 3.50**	**90.60 ± 3.69**	**89.69 ± 3.82**
DenseNet201	Naive Bayes	70.78 ± 2.11	83.49 ± 1.06	60.22 ± 4.06	62.48 ± 2.78
	MLP	76.56 ± 4.83	81.34 ± 5.81	71.74 ± 7.26	71.42 ± 6.13
	kNN	85.56 ± 4.04	86.63 ± 4.11	84.06 ± 4.61	82.99 ± 4.70
	RF	84.56 ± 4.75	86.57 ± 4.47	82.58 ± 5.79	81.44 ± 5.84
	SVM Linear	89.56 ± 2.95	89.89 ± 2.79	88.77 ± 3.34	88.29 ± 3.79
	SVM Polynomial	43.44 ± 9.11	26.70 ± 18.37	30.29 ± 4.63	50.14 ± 0.43
	SVM RBF	**90.00 ± 2.58**	**90.30 ± 2.04**	**89.25 ± 3.00**	**88.86 ± 3.60**
Inception ResNet V2	Naive Bayes	67.44 ± 5.86	65.54 ± 6.93	63.26 ± 6.85	63.18 ± 6.42
	MLP	72.22 ± 5.07	72.82 ± 7.24	68.77 ± 5.37	68.49 ± 5.22
	kNN	84.44 ± 4.22	85.55 ± 4.14	82.74 ± 5.04	81.77 ± 5.11
	RF	83.00 ± 4.25	85.07 ± 4.02	80.72 ± 5.23	79.55 ± 5.26
	SVM Linear	87.56 ± 2.52	88.16 ± 2.81	86.51 ± 2.76	85.66 ± 2.76
	SVM Polynomial	38.89 ± 0.00	19.44 ± 0.00	28.00 ± 0.00	50.00 ± 0.00
	SVM RBF	**87.56 ± 2.71**	**88.37 ± 3.38**	**86.48 ± 2.92**	**85.51 ± 2.79**
Inception V3	Naive Bayes	67.44 ± 5.94	67.49 ± 10.14	59.19 ± 8.36	60.58 ± 6.71
	MLP	74.89 ± 5.02	79.19 ± 4.33	69.57 ± 9.75	70.26 ± 7.44
	kNN	87.33 ± 4.16	89.20 ± 4.62	85.97 ± 4.59	84.70 ± 4.67
	RF	87.33 ± 3.95	90.11 ± 3.74	85.71 ± 4.64	84.18 ± 4.68
	SVM Linear	**89.56 ± 2.23**	**90.87 ± 1.93**	**88.56 ± 2.60**	**87.51 ± 2.98**
	SVM Polynomial	38.89 ± 0.00	19.44 ± 0.00	28.00 ± 0.00	50.00 ± 0.00
	SVM RBF	89.44 ± 2.24	90.87 ± 1.86	88.42 ± 2.61	87.31 ± 3.01
MobileNet	Naive Bayes	73.33 ± 4.87	74.70 ± 5.26	71.67 ± 5.27	71.60 ± 5.08
	MLP	86.89 ± 3.47	88.41 ± 2.70	86.55 ± 3.60	86.70 ± 3.44
	kNN	92.78 ± 1.88	93.01 ± 1.84	92.65 ± 1.93	92.53 ± 2.06
	RF	87.00 ± 3.62	90.03 ± 2.74	86.20 ± 3.99	85.50 ± 4.02
	SVM Linear	93.00 ± 2.33	93.16 ± 2.38	92.88 ± 2.38	92.73 ± 2.39
	SVM Polynomial	93.89 ± 2.59	94.16 ± 2.64	93.78 ± 2.64	93.60 ± 2.66
	SVM RBF	**95.33 ± 1.47**	**95.68 ± 1.31**	**95.24 ± 1.52**	**95.00 ± 1.63**
NASNetLarge	Naive Bayes	64.00 ± 2.73	69.64 ± 11.36	48.76 ± 3.90	54.60 ± 2.81
	MLP	72.56 ± 4.79	77.21 ± 6.50	67.62 ± 8.29	68.92 ± 6.62
	kNN	87.56 ± 4.30	88.54 ± 4.70	86.41 ± 4.67	85.40 ± 4.69
	RF	86.33± 2.33	88.17 ± 2.31	84.80 ± 2.82	83.52 ± 2.94
	SVM Linear	87.56 ± 3.10	88.61 ± 3.55	86.41 ± 3.59	85.45 ± 3.55
	SVM Polynomial	43.33 ± 8.89	21.67 ± 4.44	29.99 ± 3.97	50.00 ± 0.00
	SVM RBF	**90.33 ± 2.72**	**91.70 ± 3.29**	**89.47 ± 2.95**	**88.35 ± 2.88**

(Continued)

TABLE 3 | Continued

Extractors	Classifiers	Accuracy	Precision	F1 Score	Recall
	Naive Bayes	71.58 ± 3.23	71.50 ± 3.73	69.66 ± 3.51	69.39 ± 3.38
	MLP	86.74 ± 5.14	87.77 ± 4.81	86.12 ± 5.57	85.99 ± 5.46
	kNN	90.53 ± 2.58	91.35 ± 2.58	90.08 ± 2.76	89.47 ± 2.84
NASNetMobile	RF	86.00 ± 2.54	86.92 ± 2.34	85.23 ± 2.87	84.64 ± 2.92
	SVM Linear	93.89 ± 3.04	94.56 ± 2.89	93.63 ± 3.19	93.16 ± 3.42
	SVM Polynomial	57.89 ± 0.00	28.95 ± 0.00	36.67 ± 0.00	50.00 ± 0.00
	SVM RBF	**92.74 ± 2.72**	**93.32 ± 2.74**	**92.44 ± 2.87**	**91.95 ± 2.94**

The bold values are mean and standard deviation, respectively.

TABLE 4 | Continuation of **Table 3**.

Extractors	Classifiers	Accuracy	Precision	F1 Score	Recall
	Naive Bayes	63.16 ± 0.00	31.58 ± 0.00	38.71 ± 0.00	50.00 ± 0.00
	MLP	66.32 ± 1.94	82.62 ± 0.68	47.17 ± 4.62	54.29 ± 2.63
	kNN	**87.68 ± 3.40**	**89.01 ± 3.97**	**86.07 ± 3.88**	**84.71 ± 4.00**
ResNet50	RF	85.26 ± 4.16	86.81 ± 4.03	82.97 ± 5.35	81.61 ± 5.45
	SVM Linear	81.47 ± 2.50	80.59 ± 1.95	80.52 ± 2.28	81.40 ± 1.83
	SVM Polynomial	36.84 ± 0.00	18.42 ± 0.00	26.92 ± 0.00	50.00 ± 0.00
	SVM RBF	79.16 ± 1.87	77.95 ± 1.97	77.36 ± 2.15	77.25 ± 2.45
	Naive Bayes	64.00 ± 2.09	77.81 ± 4.35	51.25 ± 3.95	57.39 ± 2.46
	MLP	81.26 ± 4.34	82.14 ± 4.67	80.15 ± 4.66	79.66 ± 4.72
	kNN	90.84 ± 1.94	91.54 ± 1.89	90.44 ± 2.08	89.94 ± 2.29
VGG16	RF	87.79 ± 1.50	89.65 ± 1.70	87.00 ± 1.65	86.08 ± 1.75
	SVM Linear	86.63 ± 3.71	87.01 ± 3.67	86.16 ± 3.85	86.07 ± 3.84
	SVM Polynomial	57.89 ± 0.00	28.95 ± 0.00	36.67 ± 0.00	50.00 ± 0.00
	SVM RBF	**93.37 ± 2.45**	**94.00 ± 2.16**	**93.08 ± 2.60**	**92.64 ± 2.86**
	Naive Bayes	65.11 ± 1.59	77.94 ± 7.05	49.10 ± 2.96	55.40 ± 1.79
	MLP	78.56 ± 5.07	81.53 ± 5.42	74.76 ± 6.80	73.88 ± 6.19
	kNN	91.11 ± 2.72	91.53 ± 3.04	90.51 ± 2.81	89.97 ± 2.66
VGG19	RF	86.67 ± 3.51	88.90 ± 3.10	85.04 ± 4.27	83.69 ± 4.35
	SVM Linear	85.44 ± 3.86	84.92 ± 4.11	84.63 ± 4.04	84.56 ± 4.05
	SVM Polynomial	38.89 ± 0.00	19.44 ± 0.00	28.00 ± 0.00	50.00 ± 0.00
	SVM RBF	**91.89 ± 2.98**	**92.56 ± 2.89**	**91.24 ± 3.30**	**90.45 ± 3.51**
	Naive Bayes	66.00 ± 4.45	63.98 ± 5.23	63.02 ± 4.86	62.88 ± 4.76
	MLP	74.78 ± 4.07	75.45 ± 4.92	71.32 ± 5.30	70.95 ± 5.15
	kNN	88.11 ± 2.54	88.89 ± 3.24	87.11 ± 2.73	86.22 ± 2.72
Xception	RF	88.78 ± 2.96	90.36 ± 3.26	87.66 ± 3.32	86.45 ± 3.31
	SVM Linear	90.78 ± 2.17	91.47 ± 2.52	90.06 ± 2.35	89.23 ± 2.38
	SVM Polynomial	45.56 ± 10.18	22.78 ± 5.09	30.98 ± 4.55	50.00 ± 0.00
	SVM RBF	**92.56 ± 2.11**	**93.35 ± 2.14**	**91.97 ± 2.31**	**91.16 ± 2.49**

The bold values are mean and standard deviation, respectively. Accuracy, Precision, F1-Score, and Recall obtained through the classification of extracted features.

For the RF classifier, the criteria function was varied for Gini and entropy, the minimum number of samples that is necessary to split an internal node ranged from 1 to 6, the lowest amount of samples requested to be at a leaf node also ranged from 1 to 6, and the number of estimators was 3,000.

The training stage of the classification models considered the cross-validation technique. Of the total samples, 77 represent patients in the Alcoholic group, and 48 represent the control group. The samples were divided into ten subsets with a proportion of 80% for training and 20% for tests, randomly chosen. The hyperparameters for MLP, SVM and RF were determined through a 20-iterations random search over a cross-validation process with 10-folds.

The classification stage completes the C stage of the proposed methodology. The evaluation metrics and results are discussed below.

5.4. Evaluation Metrics

To compare our classification results with results from other methods, we use evaluation metrics based on the results obtained

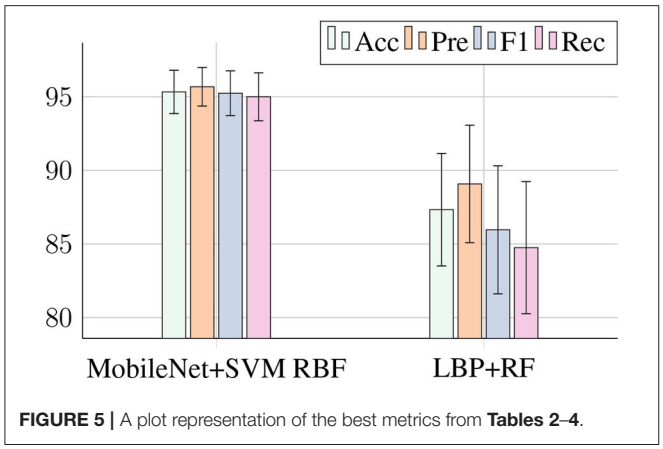

FIGURE 5 | A plot representation of the best metrics from **Tables 2–4**.

in the confusion matrix. The results of the confusion matrix include True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), some of which were used for the evaluation metrics. The evaluation of this approach used the following metrics:

Accuracy (Acc) (Fawcett, 2006) reveals the proximity of the result to the gold standard and is given by the relationship between the hits and the set of all predictions, and is presented by Equation (7).

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

The value of Precision (Fawcett, 2006) is the probability of true positives relative to all results classified as positive and is presented by Equation (8). Even if the test diagnosis is positive, this metric calculates the probability that the test will be consistent with the prior probability.

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

Recall (Rec) (Sokolova and Lapalme, 2009) represents the proportion of the results classified as positive among all the results that are really positive and is presented by Equation (9).

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

As a counterpoint to Precision, considering its risk of imbalance, the F1 Score () calculates the weighted harmonic mean between Precision and Recall and is presented by Equation (10). The F1 Score represents the performance of a method and although a diagnosis may be classified accurately, it does not mean that the method will perform the same for other data.

$$F1Score = \frac{2 * Rec * Precision}{Rec + Precision} \tag{10}$$

Except for the F1 Score index, all other evaluation measures were investigated in previous studies with signs of EEG (Ehlers et al.,

TABLE 5 | Summary and statistical comparison to other methods to classification alcoholism and healthy patients.

Author	Data	Channels	Subjects	Features/Method	Classification of features	Performance metrics			
						Acc	Precision	F1 Score	Recall
Proposed method	UCI KDD	11	20	CNN as feature extractor	MobileNet + SVM RBF	95.33 ± 1.47	95.68 ± 1.31	95.25 ± 1.52	95.00 ± 1.63
Acharya et al. (2012)	Bern Barcelona database	2	3000	Entropy, 4 HOS features, Largest Lyapunov Entropy	SVM (linear, polynomial, and RBF kernels)	91.7	93.9	–	90
Rachman et al. (2016)	UCI KDD	64	77	Daubechies wavelet family	Maximum, minimum, average and standard	85	–	–	100
Mumtaz et al. (2016)	University Malaya Medical Center	19	45	Power Spectral Density (PSD)	Logistic Regression	89.5	88.5	91	90
Ehlers et al. (1998)	University of California	1	32	CD	Discriminant analysis	88	–	–	–
Kannathal et al. (2005)	UCI KDD	60	30	CD, LLE, entropy, H	Filter by unique ranges	90	–	–	–
Faust et al. (2013)	UCI KDD	61	60	HOS cumulants	FSC	92.4	91.1	–	94.9
Patidar et al. (2017)	UCI KDD	64	122	Tunable Q-wavelet transform	Correntropy, Low-frequency(LF)-rhythms based statistical features	97.02	–	–	96.53

TABLE 6 | Accuracy, Precision, F1-Score, and Recall obtained by Acharya et al. (2012) and Mumtaz et al. (2016) proposed methods.

Method	Classifiers	Accuracy	Precision	F1 Score	Recall
Acharya et al. (2012)	**SVM Linear**	**50.67 ± 19.05**	**49.24 ± 23.82**	**46.67 ± 20.73**	**50.67 ± 19.05**
	SVM Polynomial 1	48.5 ± 11.51	46.49 ± 20.88	41.04 ± 12.04	48.5 ± 11.51
	SVM Polynomial 2	49.83 ± 4.61	44.95 ± 26.64	35.34 ± 5.66	49.83 ± 4.61
	SVM Polynomial 3	49.67 ± 2.33	34.91 ± 21.47	34.01 ± 2.69	49.67 ± 2.33
	SVM RBF	50.17 ± 2.42	37.54 ± 21.67	34.76 ± 3.50	50.17 ± 2.42
Mumtaz et al. (2016)	**Logistic Regression**	**58.00 ± 10.18**	**60.10 ± 10.64**	**54.92 ± 12.18**	**58.00 ± 10.18**

The bold values are mean and standard deviation, respectively.

1998; Acharya et al., 2012; Mumtaz et al., 2016; Rachman et al., 2016; Patidar et al., 2017).

6. RESULTS AND DISCUSSION

The proposed approach was evaluated on a computer with an Intel Core i7 microprocessor, 8 GB of RAM, a GeForce GTX 1070 Graphics Processing Unit (GPU), and a Linux LTS 16.04 operating system. The results of this paper are presented in three stages. In the first stage, the evaluation of the 21 combinations of the traditional methods for image feature extraction and classifiers. The second stage is the evaluation of the 84 combinations of CNNs as the feature extractors and classifiers. Finally, the best results are compared to related works in the last stage.

Average values and standard deviations of Accuracy, F1-Score, Precision, and Recall are shown in **Tables 2**–**4** for the features extracted with traditional methods and CNN-based methods, respectively.

Analyzing **Table 2**, GLCM-k-NN, GLCM-RF, HU-k-NN, HU-RF, LBP-k-NN, and LBP-RF stand out as they achieved at least 80% in Accuracy. Also, the RF classifier can be highlighted, since it achieved the highest Accuracy when combined with all three traditional methods. The best combination (LBP-RF) is highlighted in green. This combination reached the highest values in all four metrics.

Tables 3, **4** show the results of features extracted using CNNs, and then classified. The combinations that achieved a minimum of 90% of Accuracy and Recall were: MobileNet-k-NN, MobileNet-SVM Linear, MobileNet-SVM Polynomial, MobileNet-SVM RBF, NasNetMobile-SVM RBF, VGG16-SVM RBF, VGG19-k-NN, VGG19-SVM RBF, and Xception-SVM RBF. The combinations that had at least 90% in Accuracy, but did not achieve this value in Recall were disregarded since low values of Recall are not desirable in order not to classify alcoholism as healthy. The SVM classifier stands out when classifying deep features. This classifier obtained the best metrics values for all CNN extractors, except for ResNet50, in which the best classifier was k-NN. Among the SVM kernels, RBF reached the highest metric values for ten of the twelve CNN architectures evaluated. The best combination (MobileNet-SVM RBF) is highlighted in green.

Figure 5 compares the best combination of the traditional methods and the CNN architectures. The features extracted by the CNN-MobileNet and classified by SVM RBF achieved an accuracy 8% higher than the features extracted by the LBP and classified by RF. Also, the standard deviation for MobileNet+SVM RBF is lower, contributing to greater reliability for the system. Furthermore, even though the combination LBP+RF has an accuracy of 87%, its recall is only 84%, while the combination MobileNet+SVM RBF has accuracy and recall of 95%.

The results show that the number of features, according to **Table 1**, indicate that traditional feature extraction methods have a low representative potential. On the other hand, the feature extraction through CNN can extract more information, and this contributed to improving the classification results. Besides, tests in other bands with lower frequency channels, such as F5, TP7, PO7, and O1, did not reach metrics with values higher than 95.33%, as we achieved with the channels proposed in this work.

Acharya et al. (2012), approach, the 4 HOS features were not able to detect the most relevant features for class distinction, reaching an average accuracy of 91.7%. While the works of Ehlers et al. (1998), Kannathal et al. (2005), and Rachman et al. (2016) used statistical analysis of EEG signals. However, the use of the average value as a descriptor of the samples made the classification sensitive to extreme values. In addition, the use of descriptors with a fixed range of analysis makes it difficult to generalize unknown samples. All of these studies presented an average of <90%. **Table 5** gives a summary of the characteristics of these approaches.

The work of Faust et al. (2013) analyzed the signals using a non-linear approach. Accumulating the HOS characteristics, and combined the extractions with a Fuzzy Sugeno Classifier reached 92.4%. However, an approach using fuzzy classification imposes the need for prior knowledge of the data set for method calibration, and this makes the approach semi-automatic. Our approach does not require previous knowledge of EEG signals since the extraction models use the transfer learning techniques for feature extraction to achieve promising results.

We see in **Table 6** the results obtained by the methods proposed by Acharya et al. (2012) and Mumtaz et al. (2016). We obtain the results using extractors and classifiers proposed with the same parameters of cross-validation and dataset that we used in our method. Thus, we show the efficiency of our method within the set of EEG channels that we chose in our work. Both compared to a method that uses non-linear features and against a method that uses features in the frequency domain, respectively.

Finally, the proposed approach presented superior results to all the methods considered in this study. Our approach achieved accuracy values equivalent to the work of Mumtaz et al. (2016), considering the standard deviation. However, our approach innovated by applying a 2D analysis of the EGG signal, which allowed the application of CV techniques to overcome the problem. **Table 5** presents the results of the proposed approach of this paper compared with other works available in the literature.

7. CONCLUSIONS AND FUTURE WORKS

In this work, we proposed a new method to detect a predisposition to alcoholism from image-transformed EEG signals using traditional and deep feature extractors. We used the Learning Transfer method to extract deep image characteristics and consolidated ML methods to classify EGG signals between alcoholism and normal.

From the results presented, we can see that the CNN architectures extracted more relevant features from the samples, since the best values of Accuracy 95.33%, Precision 95.68%, F1-Score 95.24%, and Recall 95.00% were obtained in the MobileNet-SVM RBF combination. The best combination for classic extractors was LBP-RF reaching 87.33, 89.08, 85.96, and 84.75% for the same metrics.

For future work, we will apply the Principal Components Analysis (PCA) algorithm to select the most significant channels after preprocessing in order to highlight the differences between the features of each class. Another possibility is the application of fuzzy logic as a method of filtering EGG signals after preprocessing, as well as the application of mathematical morphology to highlight the differences between image textures after 1D to 2D transformation.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://kdd.ics.uci.edu/databases/eeg/eeg.data.html.

AUTHOR CONTRIBUTIONS

HZ designed and supervises all the aspects of the study implementation and drafted the manuscript. FS and AM performed the experiments based on the 2-dimensional EEG signal approach. EO applied the traditional computer vision methods for comparison with existing techniques and revised the manuscript. PR oriented the development of this work. All authors contributed to the article and approved the submitted version.

FUNDING

This study was financed in part by Tiancheng Huizhi innovation and education promotion fund of 428 the Ministry of Education (Science and technology development center of the Ministry of Education, No. 2018A010), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001. Also PPRF acknowledges the sponsorship from the Brazilian National Council for Research and Development (CNPq) via Grants Nos. 431709/2018-1 and 311973/2018-3.

REFERENCES

Acharya, U. R., Sree, S. V., Chattopadhyay, S., and Suri, J. S. (2012). Automated diagnosis of normal and alcoholic EEG signals. *Int. J. Neural Syst.* 22:1250011. doi: 10.1142/S0129065712500116

Amezquita-Sanchez, J. P., Mammone, N., Morabito, F. C., Marino, S., and Adeli, H. (2019). A novel methodology for automated differential diagnosis of mild cognitive impairment and the alzheimer's disease using EEG signals. *J. Neurosci. Methods* 322, 88–95. doi: 10.1016/j.jneumeth.2019.04.013

Begleiter, H. (2019). *EEG Database*. Available online at: https://kdd.ics.uci.edu/databases/eeg/eeg.data.html (accessed August 5, 2019).

Bhattacharyya, A., Sharma, M., Pachori, R. B., Sircar, P., and Acharya, U. R. (2018). A novel approach for automated detection of focal EEG signals using empirical wavelet transform. *Neural Comput. Appl.* 29, 47–57. doi: 10.1007/s00521-016-2646-4

Bosl, W. J., Tager-Flusberg, H., and Nelson, C. A. (2018). EEG analytics for early detection of autism spectrum disorder: a data-driven approach. *Sci. Rep*. 8:6828. doi: 10.1038/s41598-018-24318-x

Boutros, N. N., Lajiness-O'Neill, R., Zillgitt, A., Richard, A. E., and Bowyer, S. M. (2015). EEG changes associated with autistic spectrum disorders. *Neuropsychiatr. Electrophysiol*. 1:3. doi: 10.1186/s40810-014-0001-5

Breiman, L. (2001). Random forests. *Mach. Learn*. 45, 5–32. doi: 10.1023/A:1010933404324

Chollet, F. (2017). "Xception: deep learning with depthwise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu, HI), 1251–1258. doi: 10.1109/CVPR.2017.195

Chugani, H. T., Phelps, M. E., and Mazziotta, J. C. (1987). Positron emission tomography study of human brain functional development. *Ann. Neurol*. 22, 487–497. doi: 10.1002/ana.410220408

da Luz, P., and Coimbra, S. (2004). Wine, alcohol and atherosclerosis: clinical evidences and mechanisms. *Braz. J. Med. Biol. Res*. 37, 1275–1295. doi: 10.1590/S0100-879X2004000900001

da Nóbrega, R. V. M., Rebouças Filho, P. P., Rodrigues, M. B., da Silva, S. P., Júnior, C. M. D., and de Albuquerque, V. H. C. (2018). Lung nodule malignancy classification in chest computed tomography images using transfer learning and convolutional neural networks. *Neural Comput. Appl*. 32, 11065–11082. doi: 10.1007/s00521-018-3895-1

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). "Imagenet: a large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition* (Miami, FL: IEEE), 248–255. doi: 10.1109/CVPR.2009.5206848

Devor, E. J., and Cloninger, C. R. (1989). Genetics of alcoholism. *Annu. Rev. Genet*. 23, 19–36. doi: 10.1146/annurev.ge.23.120189.000315

D'Rozario, A. L., Dungan, G. C., Banks, S., Liu, P. Y., Wong, K. K., Killick, R., et al. (2015). An automated algorithm to identify and reject artefacts for quantitative EEG analysis during sleep in patients with sleep-disordered breathing. *Sleep Breath*. 19, 607–615. doi: 10.1007/s11325-014-1056-z

Ehlers, C. L., Havstad, J., Prichard, D., and Theiler, J. (1998). Low doses of ethanol reduce evidence for nonlinear structure in brain activity. *J. Neurosci*. 18, 7474–7486. doi: 10.1523/JNEUROSCI.18-18-07474.1998

Faust, O., Yanti, R., and Yu, W. (2013). Automated detection of alcohol related changes in electroencephalograph signals. *J. Med. Imaging Health Inform*. 3, 333–339. doi: 10.1166/jmihi.2013.1170

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

Foppa, M., Fuchs, F. D., and Duncan, B. B. (2001). Alcohol and atherosclerosis. *Arq. Brasil. Cardiol*. 76, 171–176. doi: 10.1590/S0066-782X2001000200009

Fukunaga, K., and Narendra, P. M. (1975). A branch and bound algorithm for computing k-nearest neighbors. *IEEE Trans. Comput. C* 24, 750–753. doi: 10.1109/T-C.1975.224297

Haralick, R. M., Shanmugam, K. S., and Dinstein, I. (1973). Textural features for image classification. *IEEE Trans. Syst. Man Cybernet.* 3, 610–621. doi: 10.1109/TSMC.1973.4309314

Haykin, S. (2008). *Neural Networks and Learning Machines*. Upper Saddle River, NJ: Prentice Hall; McMaster University.

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.

Hu, M.-K. (1962). Visual pattern recognition by moment invariants. *IRE Trans. Inform. Theory* 8, 179–187. doi: 10.1109/TIT.1962.1057692

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu, HI), 4700–4708. doi: 10.1109/CVPR.2017.243

Ibrahim, S., Djemal, R., and Alsuwailem, A. (2018). Electroencephalography (EEG) signal processing for epilepsy and autism spectrum disorder diagnosis. *Biocybernet. Biomed. Eng.* 38, 16–26. doi: 10.1016/j.bbe.2017.08.006

Jennison, K. M. (2004). The short-term effects and unintended long-term consequences of binge drinking in college: a 10-year follow-up study. *Am. J. Alcohol Abuse* 30, 659–684. doi: 10.1081/ADA-200032331

Kannathal, N., Acharya, U. R., Lim, C. M., and Sadasivan, P. (2005). Characterization of EEG–a comparative study. *Comput. Methods Programs Biomed.* 80, 17–23. doi: 10.1016/j.cmpb.2005.06.005

Koley, B., and Dey, D. (2012). An ensemble system for automatic sleep stage classification using single channel EEG signal. *Comput. Biol. Med.* 42, 1186–1195. doi: 10.1016/j.compbiomed.2012.09.012

Kozel, F. A., Revell, L. J., Lorberbaum, J. P., Shastri, A., Elhai, J. D., Horner, M. D., et al. (2004). A pilot study of functional magnetic resonance imaging brain correlates of deception in healthy young men. *J. Neuropsychiatry Clin. Neurosci.* 16, 295–305. doi: 10.1176/jnp.16.3.295

Längkvist, M., Karlsson, L., and Loutfi, A. (2014). A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recogn. Lett.* 42, 11–24. doi: 10.1016/j.patrec.2014.01.008

Lloyd-Fox, S., Blasi, A., and Elwell, C. (2010). Illuminating the developing brain: the past, present and future of functional near infrared spectroscopy. *Neurosci. Biobehav. Rev.* 34, 269–284. doi: 10.1016/j.neubiorev.2009.07.008

McBride, J. C., Zhao, X., Munro, N. B., Jicha, G. A., Schmitt, F. A., Kryscio, R. J., et al. (2015). Sugihara causality analysis of scalp EEG for detection of early Alzheimer's disease. *Neuroimage Clin.* 7, 258–265. doi: 10.1016/j.nicl.2014.12.005

Mohammadi, M. R., Khaleghi, A., Nasrabadi, A. M., Rafieivand, S., Begol, M., and Zarafshan, H. (2016). EEG classification of adhd and normal children using non-linear features and neural network. *Biomed. Eng. Lett.* 6, 66–73. doi: 10.1007/s13534-016-0218-2

Mumtaz, W., Vuong, P. L., Xia, L., Malik, A. S., and Rashid, R. B. A. (2016). Automatic diagnosis of alcohol use disorder using EEG features. *Knowl. Based Syst.* 105, 48–59. doi: 10.1016/j.knosys.2016.04.026

Muñoz-Organero, M., Powell, L., Heller, B., Harpin, V., and Parker, J. (2018). Automatic extraction and detection of characteristic movement patterns in children with adhd based on a convolutional neural network (cnn) and acceleration images. *Sensors* 18:3924. doi: 10.3390/s18113924

Ojala, T., Pietikainen, M., and Harwood, D. (1994). "Performance evaluation of texture measures with classification based on kullback discrimination of distributions," in *Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol. 1-Conference A: Computer Vision & Image Processing* (Jerusalem: IEEE), 582–585. doi: 10.1109/ICPR.1994.576366

Ojala, T., Pietikäinen, M., and Mäenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* 24, 971–987. doi: 10.1109/TPAMI.2002.1017623

Patidar, S., Pachori, R. B., Upadhyay, A., and Acharya, U. R. (2017). An integrated alcoholic index using tunable-q wavelet transform based features extracted from EEG signals for diagnosis of alcoholism. *Appl. Soft Comput.* 50, 71–78. doi: 10.1016/j.asoc.2016.11.002

Rachman, N. T., Tjandrasa, H., and Fatichah, C. (2016). "Alcoholism classification based on EEG data using independent component analysis (ICA), wavelet de-noising and probabilistic neural network (PNN)," in *2016 International Seminar on Intelligent Technology and Its Applications (ISITIA)* (Lombok: IEEE), 17–20. doi: 10.1109/ISITIA.2016.7828626

Ren, G.-P., Yan, J.-Q., Yu, Z.-X., Wang, D., Li, X.-N., Mei, S.-S., et al. (2018). Automated detector of high frequency oscillations in epilepsy based on maximum distributed peak points. *Int. J. Neural Syst.* 28:1750029. doi: 10.1142/S0129065717500290

Rodrigues, J., d,. C., Rebouças Filho, P. P., Peixoto Jr, E., Kumar, A., and de Albuquerque, V. H. C. (2019). Classification of EEG signals to detect alcoholism using machine learning techniques. *Pattern Recogn. Lett.* 125, 140–149. doi: 10.1016/j.patrec.2019.04.019

Rundo, F., Rinella, S., Massimino, S., Coco, M., Fallica, G., Parenti, R., et al. (2019). An innovative deep learning algorithm for drowsiness detection from EEG signal. *Computation* 7:13. doi: 10.3390/computation7010013

Sasaki, Y. (2007). The truth of the f-measure. *Teach. Tutor Mater.* 1, 1–5.

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arxiv [Preprint] arXiv*:1409.1556.

Snodgrass, J. G., and Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *J. Exp. Psychol. Hum. Learn. Mem.* 6:174. doi: 10.1037/0278-7393.6.2.174

Sokolova, M., and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Inform. Process. Manage.* 45, 427–437. doi: 10.1016/j.ipm.2009.03.002

Stam, C. (2010). Use of magnetoencephalography (MEG) to study functional brain networks in neurodegenerative disorders. *J. Neurol. Sci.* 289, 128–134. doi: 10.1016/j.jns.2009.08.028

Suykens, J. A., and Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Process. Lett.* 9, 293–300. doi: 10.1023/A:1018628609742

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Boston, MA), 1–9. doi: 10.1109/CVPR.2015.7298594

Theodoridis, S., and Koutroumbas, K. (2008). *Pattern Recognition. 4th Edn.* Burlington, MA: Academic Press.

Tzimourta, K. D., Giannakeas, N., Tzallas, A. T., Astrakas, L. G., Afrantou, T., Ioannidis, P., et al. (2019). EEG window length evaluation for the detection of Alzheimer's disease over different brain regions. *Brain Sci.* 9:81. doi: 10.3390/brainsci9040081

Vapnik, V. N. (1998). *Statistical Learning Theory*. Red Bank, NJ: John Wiley & Sons; EUA.

Wang, L.-J., Wu, C.-C., Lee, M.-J., Chou, M.-C., Lee, S.-Y., and Chou, W.-J. (2019). Peripheral brain-derived neurotrophic factor and contactin-1 levels in patients with attention-deficit/hyperactivity disorder. *J. Clin. Med.* 8:1366. doi: 10.3390/jcm8091366

World Health Organization (2019). *Global Status Report on Alcohol and Health 2018*. Geneva: WHO Press.

Wu, S., Zhong, S., and Liu, Y. (2018). Deep residual learning for image steganalysis. *Multimed. Tools Appl.* 77, 10437–10453. doi: 10.1007/s11042-017-4440-4

Zhang, X. L., Begleiter, H., Porjesz, B., Wang, W., and Litke, A. (1995). Event related potentials during object recognition tasks. *Brain Res. Bull.* 38, 531–538. doi: 10.1016/0361-9230(95)02023-5

Zoph, B., and Le, Q. V. (2016). Neural architecture search with reinforcement learning. *arxiv [Preprint] arXiv*:1611.01578.

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

frontiers in Human Neuroscience

REVIEW
published: 09 February 2021
doi: 10.3389/fnhum.2021.638887

Brain-Machine Interfaces to Assist the Blind

Maurice Ptito[1,2,3], Maxime Bleau[1], Ismaël Djerourou[1], Samuel Paré[1], Fabien C. Schneider[4,5] and Daniel-Robert Chebat[6,7]*

[1] École d'Optométrie, Université de Montréal, Montréal, QC, Canada, [2] Department of Nuclear Medicine, University of Southern Denmark, Odense, Denmark, [3] Department of Neuroscience, University of Copenhagen, Copenhagen, Denmark, [4] TAPE EA7423 University of Lyon-Saint Etienne, Saint Etienne, France, [5] Neuroradiology Unit, University Hospital of Saint-Etienne, Saint-Etienne, France, [6] Visual and Cognitive Neuroscience Laboratory (VCN Lab), Department of Psychology, Faculty of Social Sciences and Humanities, Ariel University, Ariel, Israël, [7] Navigation and Accessibility Research Center of Ariel University (NARCA), Ariel, Israël

The loss or absence of vision is probably one of the most incapacitating events that can befall a human being. The importance of vision for humans is also reflected in brain anatomy as approximately one third of the human brain is devoted to vision. It is therefore unsurprising that throughout history many attempts have been undertaken to develop devices aiming at substituting for a missing visual capacity. In this review, we present two concepts that have been prevalent over the last two decades. The first concept is sensory substitution, which refers to the use of another sensory modality to perform a task that is normally primarily sub-served by the lost sense. The second concept is cross-modal plasticity, which occurs when loss of input in one sensory modality leads to reorganization in brain representation of other sensory modalities. Both phenomena are training-dependent. We also briefly describe the history of blindness from ancient times to modernity, and then proceed to address the *means* that have been used to help blind individuals, with an emphasis on modern technologies, invasive (various type of surgical implants) and non-invasive devices. With the advent of brain imaging, it has become possible to peer into the neural substrates of sensory substitution and highlight the magnitude of the plastic processes that lead to a rewired brain. Finally, we will address the important question of the value and practicality of the available technologies and future directions.

Keywords: blindness, cross-modal plasticity, sensory substitution device, visual prostheses, sensory substitution

OPEN ACCESS

Edited by:
Marcin Wozniak, Silesian University of Technology, Poland

Reviewed by:
Kevin C. Chan, New York University, United States
Martin Lages, University of Glasgow, United Kingdom

****Correspondence:***
Maurice Ptito
maurice.ptito@umontreal.ca; ptito.maurice@gmail.com

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

Received: *07 December 2020*
Accepted: *19 January 2021*
Published: *09 February 2021*

Citation:
Ptito M, Bleau M, Djerourou I, Paré S, Schneider FC and Chebat D-R (2021) Brain-Machine Interfaces to Assist the Blind. Front. Hum. Neurosci. 15:638887. doi: 10.3389/fnhum.2021.638887

HISTORY OF BLINDNESS

For most sighted people, the very thought of blindness awakens a deep fear: a fear of the unknown, of an "endless night," of being unable to move and orient oneself (Commend, 2001). This fear has had repercussions throughout recorded history and on the conditions of people living with blindness.

A Limiting Vision of Blindness: From Ancient World to Enlightenment

Throughout the ages, blindness has long been associated with mythical or biblical beliefs to provide lessons or even to give inspiration to the "common people." In Ancient Greece, blindness was

generally viewed as a punishment from the Gods. Indeed, although Homer was rumored to be blind, the scarce reports that remain of this period depict blindness as being associated with accidents, war injuries and, importantly, punishment for transgressions (Barasch, 2001). That preconception persisted through the Middle Ages when blindness and other disabilities were often viewed as acts of god and deliberate blinding was the most dreaded of punishments (Wheatley, 2010). People living with blindness were thus associated with misery and were often depicted as beggars or as praying for a miracle of the sort attributed to Jesus (Weygand, 2009). Because of this prevailing attitude toward blindness, blind people long found themselves objects of derision or charity, whose existence was often reduced to their reliance on the help of others for daily living (Barasch, 2001; Weygand, 2009; Wheatley, 2010; Margo et al., 2013). This view role, however, began to change and improve in Europe during the Enlightenment of the 17th and 18th centuries. The separation between blindness and biblical beliefs found first expression in William Molyneux's question addressed in 1688 to John Locke, cited in *An Essay Concerning Humane Understanding*:

> *"A Man, being born blind and having a Globe and a Cube [. . .], Let us suppose his Sight Restored to Him; Whether he Could, by his Sight, and before he touch them, know which is the Globe and which is the Cube?"* [from Ferretti and Glenney (2020)].

The question was later entertained by other early modern philosophers such as Gottfried Leibniz, George Berkeley, Adam Smith, and many others. While this purely philosophical question did not directly address the inclusion or education of the blind, it allowed further conjectures about perceptual learning, multi-sensory integration and the capacity of the blind to learn without the use of vision (Ferretti and Glenney, 2020).

Education Through Touch: From Diderot to Braille and Howe

As education and writing assumed greater importance during the Enlightenment, there arose many examples of blind individuals who successfully educated themselves and accomplished inspirational feats. Notably among them, Nicholas Saunderson (1682–1739), a scholar at the University of Cambridge, became a tutor in mathematics and physics and won the esteem of Newton himself who judged him one of his few contemporaries who truly understood the value of his work. There also were Mélanie de Salignac (1744–1766), a musician who learnt by herself how to read, write and correspond with friends using cutout letters, and Maria Theresia von Paradis (1759–1824), who was a talented singer and pianist (see **Figure 1**). Such individuals became sources of inspiration for Denis Diderot (1713–1784) in the writing of his 1749 essay *The Letter on the Blind for the Benefit of Those Who See*, where he lauded the abilities of blind people. According to Diderot, educating the blind in writing and reading was possible through the sense of touch (Margo et al., 2013). His philosophy offered a foundation for the efforts toward the education of blind people in the centuries that followed, being one of the first savants who truly focused on their ability rather than disability. Indeed, Diderot's philosophy was central to Valentin Haüy's work in founding the first school for the blind in 1784 (now known as the *Institut national des jeunes aveugles* or INJA). Valentin Haüy (1745–1822) was a French calligraphy professor who proved that blind individuals could learn to read embossed text with the use of their fingers. He invented the first reading system of raised Roman letters which he successfully taught for years. Haüy's school later gave birth to the Braille alphabet, a new tactile writing system invented by one of its blind students: Louis Braille [reviewed in Jiménez et al. (2009)]. Louis Braille (1809–1852) was inspired by the Night Writing (from French: *écriture nocturne*) system previously developed by Charles Barbier de la Serre for the use of French soldiers who had to read and write in the dark while on campaign. Barbier's system was based on phonetics and consisted of different combinations of raised points on a two by six grid of twelve points. This concept was deemed too cumbersome by Braille, who went to create a two by three grid of six points representing the alphabetical system that was simpler and easier to learn. In 1829, then 15-year-old Louis Braille published his first version of the system, which was officially adopted in the school and in France in 1854. The eponymous Braille system was the first successful sensory substitute for reading without vision and it is still in wide use today. In fact, Braille and the capacity to read through touch constituted a colossal step forward for the rehabilitation of blind people in society, a concept that was promoted abroad by 19th century reformers such as Samuel Gridley Howe, who founded the New England Institution for the Education of the Blind (now the Perkins School for the Blind). **Figure 2** illustrates the development stages of the embossed letter system. It is now fully appreciated that blind people can be trained to substitute their intact senses for vision, enabling them to become integral, productive and autonomous members of society. Indeed, the blind can even develop supra-normal sensory abilities through the overtraining of other modalities.

Understanding Echolocation, the "Sixth Sense" of the Blind

Supra-normal abilities of blind people in other sensory modalities such as touch and audition are well known today, but were first reported as soon as 1749 in Diderot's work cited above. Indeed, Diderot was among the first to report the blind's use of echolocation or, as he discussed, their ability to perceive objects and estimate their distance *via* sensations manifesting on perceived on the skin of the face. Diderot attributed this phenomenon to the compression of air against the skin upon approaching an object. According to Diderot, the facial nerves and sensory end organs had increased sensitivity in the blind. Thus, for many years it was held that the blind could feel changes in air pressure with their forehead and cheeks (Burklen, 1924), an ability that was named "*perceptio facialis*," or "facial vision" (Levy, 1872). At the start of the 20th century, authors began to debate the nature of "facial vision" and on whether it was due to the use of reflected sounds (Dresslar, 1893; Heller, 1904; Truschel, 1906; Villey, 1930), air pressure (James, 1890), "ether waves" (Javal, 1905) or even "vestigial Ranvier corpuscles" in the skin of the forehead (Romains, 1924). This debate continued

FIGURE 1 | Important individuals in the history of blindness. **(A)** Portrait of John Locke and cover and first page of his paper: *An Essay on Humane Understanding*; **(B)** Portrait of Denis Diderot and cover and first page of his article: *A Letter on the Blind for the Use of Those Who See*; **(C)** Portrait of Nicholas Saunderson; **(D)** Drawing of Maria Theresia Von Paradis; **(E)** Portrait of Louis Braille; and **(F)** Picture of Russell Williams (photo courtesy AER O&M Division Warren Bledsoe Archives, American Printing House for the Blind).

until the period following World War II, and the conducting of the *Cornell Experiments*, a series of notable experiments where numerous authors systematically investigated the nature of "facial vision" [reviewed in Thaler and Goodale (2016)]. It was then discovered that "facial vision" was not based on atmospheric pressure cues felt on the face but rather that this skill was attributable to the use of auditory cues. When the ears of blind participants were plugged, they were no longer able to use "facial vision" (Supa et al., 1944). This generation of authors understood that blind individuals were using a form of echolocation (as in bats and dolphins) to perceive reflected sounds, sound shadows and changes in the sound (i.e., the Doppler effect) in manners unavailable to sighted people (Supa et al., 1944; Worchel and Dallenbach, 1947; Worchel et al., 1950). It was subsequently concluded that these auditory abilities were however, cross-modally experienced as tactile sensations of pressure against the face (Kohler, 1964), at the conclusion of the long-lasting debate on the nature of "facial vision." The capacity for echolocation was found to be present in 85% of blind individuals and to correlate with the age of blindness onset and its duration (Juurmaa, 1965). Congenitally blind individuals (CB) proved to be more effective in the use of echolocation than their sighted counterparts (Supa et al., 1944; Juurmaa, 1965; Strelow and Brabyn, 1982; Boehm, 1986). However, it was soon established that blindfolded sighted individuals could learn the skill of echolocation as could individuals who acquired their blindness later in life (Worchel and Ammons, 1953). Thus, "facial vision" (properly echolocation) was and is still viewed as an essential skill for the blind to learn to achieve a higher level of independence. Indeed, in today's orientation and mobility (O&M) training, blind individuals are taught to use echolocation and environmental sounds in conjunction with the white cane and other technology to navigate safely and independently.

Toward Independent Travel: The White Cane and O&M Training

O&M training, as we know it today, is still a developing field that traces its roots to World War II (Sauerberger, 1996; Bledsoe, 2010). During those years, blind people were taught to use "facial vision" and other orientation strategies (i.e., memorizing lay-outs and landmarks) with instruction from "orientors" in rehabilitation programs. However, this approach was often prioritized over the cane and other tools that could contribute to the perceived stigma of blindness (Bledsoe, 2010). The numerous American soldiers blinded due to the vicissitudes of war were sent to military hospitals such as Valley Forge and Dibble, where they healed from their wounds and learned to navigate with a cane before being transferred to the rehabilitation program in Avon, Ohio. In order to treat the growing numbers of injured and visually impaired, Valley Forge hired Richard E. Hoover and Warren Bledsoe who had previously worked as teachers

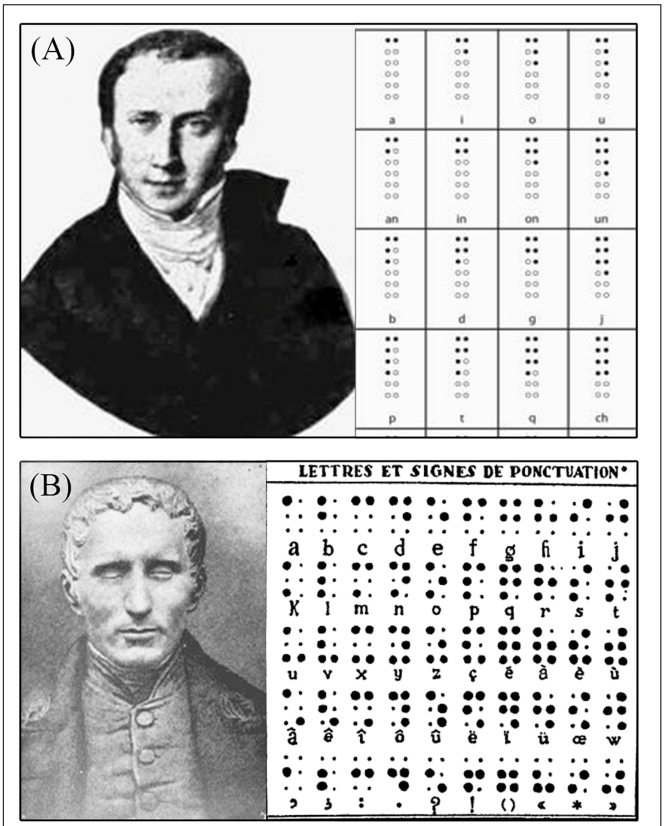

FIGURE 2 | The creation of Braille. **(A)** Portrait of Charles Barbier and examples of his Night Writing system based on phonetics, which is the system that inspired Louis Braille; **(B)** Portrait of Louis Braille and the early version of the Braille writing system.

for blind individuals. Early in their postings at Valley Forge, they quickly concluded that echolocation alone was insufficient to support safe and efficient navigation during which obstacle avoidance was necessary.

> *"[. . .] the first thing they need is to know how to get around. We've been working on it, but not enough. [. . .] People say blind people in this country do a good job of getting around. I don't think they do a good job. I think they do a hell of a poor job."* - Richard Hoover [from Bledsoe (2010)].

While we know that canes and staffs have been used for millennia, as attested by numerous examples from ancient Greece and biblical texts (Levy, 1872), the internationally recognized white cane was invented in 1921 and was first promoted by the Lions Club International in 1931. However, Hoover promoted the use of a more lightweight cane and developed the foundations of cane techniques as taught today. Indeed, Hoover blindfolded himself and, alongside Bledsoe and other instructors, experimented with new and affective cane techniques. Based on this experience, he established the "touch cane technique" and trained other instructors in its proper use for the benefit of blinded soldiers. One of those soldiers was Russel Williams, who had lost his sight from injuries during the Normandy invasion. Williams later transferred to the program in Avon, Ohio, where he learned echolocation and orientation techniques, and decided on his own accord to merge all his training to achieve greater autonomy. In 1948, he was appointed as the chief of the new rehabilitation program at the Veterans Administration Hospital in Hines, Illinois. In that time, he worked alongside Bledsoe to enroll and train new specialists in the field of "foot travel," which later became known as O&M training (Sauerberger, 1996; Bledsoe, 2010). While World War II had disastrous repercussions on the world, it enabled the initiation of greatly improved rehabilitation services offered to individuals with vision impairments, which remain in use to this day. Proceeding from the experiments on echolocation to the development of the touch cane technique, the work following the war enabled the growth of O&M training which now plays a pivotal role in the rehabilitation of visually impaired individuals around the world seeking greater autonomy, confidence, and a better quality of life.

Modern Technologies: Brain Interfaces to Help the Blind "See"

Technologies and tools introduced in O&M training help blind individuals to expand their perception of the environment and thus extend their domains of action. To date, the white cane remains the main compensatory tool utilized by blind individuals worldwide. As an extension of the arm, the white cane provides safety against obstacles by extending the range of detection and provides additional information (auditory and tactile) on the environment such as changes in floor texture and denivelation [reviewed in Guth et al. (2010)]. However, the white cane, even when used in conjunction with echolocation, has a significant limitation; While it detects objects on the ground, the upper body and head remain unprotected and blind individuals are still at risk of dangerous collisions that they cannot anticipate (Suterko, 1967). Consequently, the blind suffer disproportionately more injuries due to collisions to the head, and are likewise vulnerable to the risk of falls (Manduchi and Kurniawan, 2011) which can contribute to the feeling of anxiety about travels and, ultimately, lead to social isolation (Beggs, 1992). Faced with this issue, it is not surprising that many scientists began working on new technologies aiming to enhance the corporal safety of individuals living with blindness, most ambitiously in efforts to restore sight. These efforts employ brain interface technologies that can tap into the visually deprived brain's potential of adaptation to new stimuli and tasks. These new brain interfaces can expand the perception of the blind beyond the capacities of the white cane and Braille, thus affording more opportunities to learn, travel safely, and participate as independent members of society.

Today, there are many kinds of brain interfaces aiming to help the blind "see," which we classified into two main categories of devices. The first category consists of *invasive brain interfaces* that require surgical implementation of the device in the visual system, such as retinal and cortical implants, in order to restore sight to those who lost it. The second category consists of *non-invasive brain interfaces*, such as electronic and electromechanical aids, that aim to complement the sensory abilities the blind

already possess, along with sensory substitution devices (SSDs), which aim to offer a "visual-like" experience by electronically translating visual information into another modality, such as touch and audition.

BRAIN INTERFACES FOR VISION RECOVERY

Invasive Brain Interfaces

Researchers have tested many brain sites for electrical stimulation aiming to restore vision with the use of implants. For example, there have been attempts to electrically stimulate the visual cortex and retinal cells of patients with blindness due to retinitis pigmentosa (RP) or age-related macular degeneration (AMD). In addition, other studies have shown that it is possible to evoke sensory perceptions by stimulating the optic nerve (Delbeke et al., 2001) or the lateral geniculate nucleus (Pezaris and Reid, 2007) although these techniques are not widely used because of the many challenges and risks associated with the neurosurgical procedures to access such inner brain structures (Allen and Ayton, 2020).

Cortical Visual Prostheses

Restoring vision has been of interest to scientists for several centuries. Charles Le Roy, a French physicist, was interested in curing diseases with electricity. In an attempt to cure a patient of blindness, he developed a metal device that applied to the head of the patient and connected it to a Leyden jar. Surprisingly for the time, the patient reported perceiving flashes of light during the electric shocks (LeRoy, 1755). This was the first recorded demonstration of the electrical excitability of the visual cortex, and was the inspiration of a series of attempts for vision recovery. In the early 20th century, neurosurgeons made use of the research opportunity presented by awake opened skull patients to electrically stimulate their visual cortex, which evoked the experience of retinotopically organized phosphenes. The spatial representation of the visual field in the human primary visual cortex was discovered using these techniques (Holmes, 1918; Löwenstein and Borchardt, 1918). This approach later prompted John C. Button to develop a device aiming to restore vision to blind people by electrical stimulation of the occipital cortex. In a test of the device, a blind patient reported seeing flashes of light and was able to locate and assess the brightness of a light source (Button, 1958). Some years later, Brindley and Lewin (1968) produced a wireless prototype of a cortical visual prosthesis, which consisted of 80 extracranial radio receivers connected to 80 intracranial electrodes inserted inside the calcarine fissure. The prototype did not support reading as the authors had hoped, but did allow simple pattern discrimination. At around the same time, William Dobelle developed a removable visual neuroprosthesis that allowed him to stimulate the visual cortex of patients undergoing brain surgery [reviewed in Lewis et al. (2015)]. These pioneering studies set the stage for the development of more sophisticated instrumentations and new generation of cortical implants. In 2020, several projects are in progress and clinical trials are underway or planned in the coming years (for a review on neurobionics and cortical implants see: Allen and Ayton, 2020; Chen et al., 2020; on retinal implants see: Nowik et al., 2020).

CORTIVIS

The aim of the CORTIVIS project is to capture the visual scene using a bioinspired artificial retina designed to emulate aspects of the visual processing that occur in the retina. The CORTIVIS project uses the Utah Electrode Array (UEA), which consists of 100 electrodes of 1.0–1.5 mm in length. It is designed to reach the cortical layer 4c (the target of geniculate innervations) and to limit damage to neurons. Early experiments showed that the electrical stimulation of the implanted electrodes elicited visual perception in monkeys (Normann et al., 2009) and preliminary investigations were carried out in human patients with epilepsy or brain tumors during brain surgery. Promising results were obtained with safe implantation, high-quality visual cortex recordings and induced perception of phosphenes (Fernandez et al., 2015). Recently, a new system coined "The High-Channel-Count Neuroprosthesis" has been successfully tested on monkeys. It uses a high number of implanted electrodes (1,024 in total) placed in the geniculate recipient layer of the primary visual cortex (V1) and in area V4 of the ventral visual stream. Monkeys equipped with such implants were able to recognize simple shapes, motion and letters (Chen et al., 2020).

Orion

This system consists of a camera, a computer and a subdural array of 60 surface electrodes applied to the medial occipital lobe. After processing of the video image, the information is transmitted wirelessly to the array. A preliminary study in one blind patient demonstrated the safety and basic functional aspects of the device. Ongoing clinical trials that started in late 2017 have so far included five blind patients with a follow-up planned for 5 years (Niketeghad and Pouratian, 2019). Preliminary results indicated that patients were able to perceive phosphenes (Barry et al., 2020).

ICPV Project

The Intracortical Visual Prosthesis Project (ICVP) uses a Wireless Floating Microelectrode Array (WFMA) consisting of 16 parylene-insulated iridium microelectrodes placed on the surface of the visual cortex, an integrated circuit microprocessor and a microcoil with wireless power and activation. A video camera mounted on eyeglasses or a headband connects to the video processor unit that converts images into a pattern that maps to the array of electrodes. The signal is then transmitted to the telemetry controller located on the head *via* the stimulation modules that distribute signals and power wirelessly to each WFMA module. Human clinical trials are ongoing (Troyk, 2017).

Gennaris

This setup consists of a camera mounted on eyeglasses to capture the scene and transmit it to a "Pocket Processor" that extracts useful information and then sends it to the tiles (43 intracortical electrodes per tiles) implanted in layer 4 of the primary visual cortex. Signals are broadcast by a wireless transmitter located at the back of the head (Lowery et al.,

2015). Safety tests on experimental animals have confirmed the production of phosphenes, and histological examination reveals minimal damage to the cortex after implantation and that long-term stimulation is possible without adverse events (Lowery et al., 2017; Rosenfeld et al., 2020). The first human clinical trials are planned in the coming years.

Retinal Implants

Retinal prostheses have been developed as potential treatments for retinal pathologies such as RP and AMD, which are the leading causes of blindness. In these pathologies, the retinal ganglion cell (RGC) layer is relatively unaffected, making such patients good candidates for intraretinal implantation (Santos et al., 1997; Medeiros and Curcio, 2001). Retinal prostheses are classified according to the locus of the electrode array, i.e., epiretinal, subretinal, and suprachoroidal.

Argus Retinal Prostheses

The first retinal prosthesis was the Argus® I, which is an epiretinal array of 16 electrodes wirelessly connected to a computer and a camera. Clinical trials with the implant indicated that patients were able to accomplish simple visual detection and discrimination tasks (Yanai et al., 2007). However, the spatial resolution was inherently limited by the number of electrodes and the distance between them (Caspi et al., 2009). To overcome such limitations, the subsequently the Argus® II that boasts an epiretinal array of 60 (6×10) platinum electrodes and better spatial resolution of the transmitted signal. Implanted patients were able to discriminate and recognize 2D and 3D objects, identify large high contrast letters (Stronks and Dagnelie, 2014) localize targets (Ahuja et al., 2011) and detect motion (Arsiero et al., 2011; Luo et al., 2014). Moreover, in a simple navigation task, patients were able to follow a high contrast line on the ground and find a door (Humayun et al., 2012).

Alpha-IMS

Alpha-IMS is a subretinal implant placed in an area devoid of photoreceptors with the goal to act as a substitute for the missing photoreceptors. It consists of a chip composed of 1,500 photodiodes that detect light, an amplifier circuit and penetrating electrodes. The amplified signal activates the bipolar cells (Stingl et al., 2013). Using this implant, patients were able to perceive and localize a light source, and detect motion. The second-generation of the device, the Retina Implant Alpha AMS is an improved version with 1,600 photodiodes and increased durability, and is now being tested (Edwards et al., 2018).

The Bionic Vision Australia

The Bionic Vision Australia (BVA) is a suprachoroidal implant that reduces the surgical risks of causing damage to the retina. Since the implant is far from the targeted retinal cells, patients demonstrated very poor visual acuity (20/8397) with the device (Ayton et al., 2014).

Figure 3 illustrates four types of invasive implants (retinal, in the optic nerve, thalamic and cortical).

Non-invasive Brain Interfaces Through Touch and Audition

While invasive interfaces require decisive surgery and have not proven their efficacy, new attempts have been made in developing non-invasive devices. Since the beginning of the 20th century, researchers have developed sensory substitution systems to replace vision with other senses like touch and audition.

Electronic Aids for Reading

As described in section "Education Through Touch: From Diderot to Braille and Howe," Braille brings to blind people a universal writing and reading system. However, the blind people must still rely upon sighted persons to translate printed texts into Braille, or to provide audio transcripts. Several devices have been designed to give the blind broader independence in reading.

The Optophone, developed in 1912, was one of the first sensory aid systems to transduce light into sound. First designed for enabling independent mobility, it later found application as a reading aid. Equipped with the device, which applies mechanical signals to the hand, some blind people were able to read at a rate up to 60 words per minute (d'Albe, 1920). This early success inspired some scientists to consider the incredible potential of the tactile sense for sending "visual" information to the brain. Indeed, Geldard (1957) developed a vibrotactile device based on a communication code like Morse code that could transmit individual letters to the reader (Geldard, 1957). Bliss et al. (1963) took the idea one step further by using air puffs to the chest as tactile stimulators, and found that (with training) blind subjects could perceive apparent motion with good spatial and temporal acuity (Bliss et al., 1963). In 1966, Bliss went on to design a system of vibrotactile stimulators consisting of 96 piezoelectric pins, each connected to photocells, which enabled the blind to perceive printed texts. By placing their index finger on the piezoelectric grid, users could feel the vibrations corresponding spatially to the letters. After 50 h of practice, some participants could read at a speed of 30 words per minute, thus one third of the rate for skilled Braille readers (Linvill and Bliss, 1966). This system became commercially available in the 1970s under the name Optacon (Optical to Tactile Conversion), but did no longer find great success in the 1990s since it was surpassed by the advent of scanners equipped with optical character recognition software that became generally less expensive, easier and faster to use to access printed literature without vision (Stein, 1998; Kendrick, 2005).

Electronic Travel Aids to Assist Mobility

To improve personal safety during navigation, electronic travel aids (ETAs) mainly function on the echo principle of active energy-radiating systems. Indeed, most ETAs are devices that detect obstacles by emitting a form of energy and capturing its reflection with a sensor. ETAs can deliver to the blind user information about looming obstacles, communicated by easily understandable auditory or tactile stimulations.

Electromagnetic Radiation

Electronic travel aid devices working on the emission of electromagnetic radiation (light), often functioned through

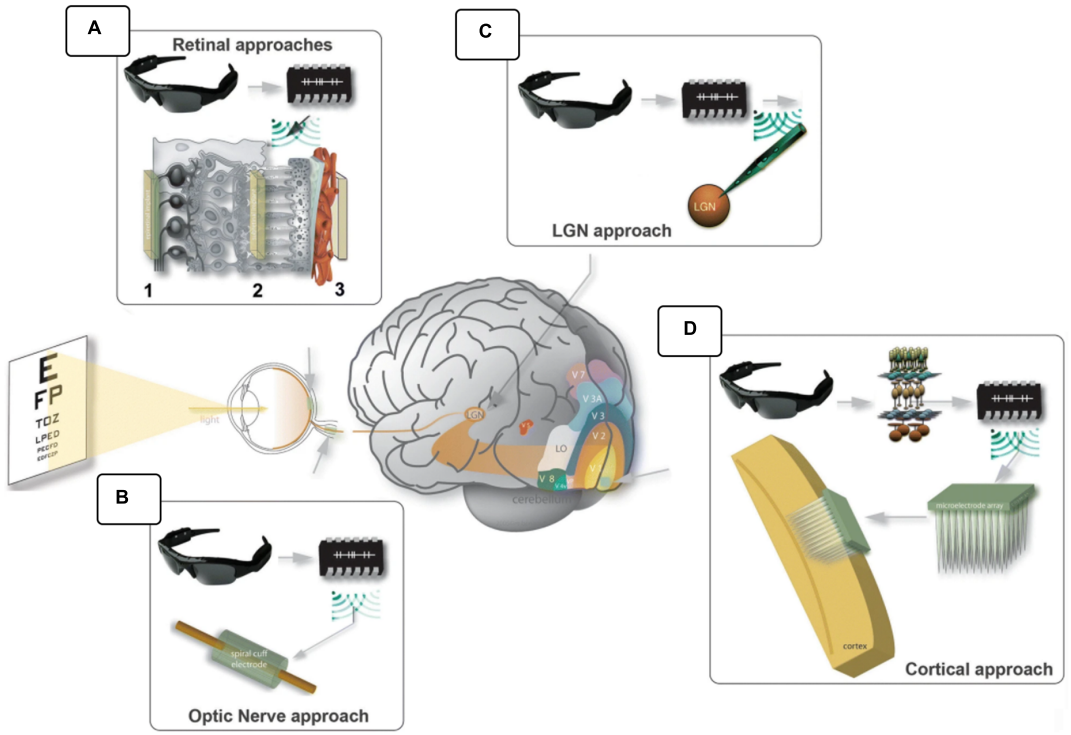

FIGURE 3 | Visual prostheses. In general, the scene is captured by a video camera, processed by a computer unit and sent to the electrical interface that stimulates the visual pathways. Different anatomical locations were explored: **(A)** epiretinal (1), subretinal (2) and suprachoroidal (3); **(B)** the optic nerve; **(C)** the lateral geniculate nucleus; and **(D)** the visual cortex. (modified from Fernandez, 2018; https://creativecommons.org/licenses/by/4.0/).

optical triangulation (Benjamin, 1974). The light rays reflected by the tangible surface (or obstacle) enter the sensor (photodetector) at various angles depending on the object's distance. The incident angle thus encodes distance information. The first effective ETAs, known as Obstacle Detectors, enabled the detection of objects by sending a single beam of light from a hand-held flashlight-like source. They signaled the detected obstacle with a vibration of the handle, thus permitting the users to detect and avoid obstacles in various environments (maze, street, store) but users proved to be slower than with their habitual white cane (Benjamin, 1968). Furthermore, these devices could not detect changes in floor texture or elevation and participants, thus, preferred using them in combination with their cane. This finding led to the development of a system combining the cane and the light beam, the laser cane, which was equipped with three laser sources pointing at different angles (downward, forward, and upward), thus aiming to extend the range of the cane while enabling the detection of higher obstacles (Benjamin, 1974). The latest prototype, the laser cane N-2000, was used in the 2000s, but is no longer available because it was significantly more expensive than similar ultrasonic ETAs (Roentgen et al., 2008; Li, 2015).

Scanning With Ultrasounds

Ultrasonic signals have a slower propagation speed than light, which naturally leads to longer reflection delays, allowing for more precise measurements of distances compared to optical triangulation (National Research Council, 1986). For this reason, contemporary ETAs still use ultrasounds. One of the first successful ultrasonic ETAs was Russell's Travel Pathsounder, a pendant-like device that emitted a conical ultrasonic beam for obstacle detection. It reduced collision risks by signaling obstacles in the immediate navigational environment with simple sounds and vibrations as warnings (Russell, 1967). A later device, the Sonic Guide (successor to the Sonic Torch and Binaural Sensory Aids), enabled some degree of object discrimination and localization with more complex auditory cues (Kay, 1964). The Sonic Guide technology became the foundation of the "K"-Sonar (Penrod et al., 2009), a smaller compact sensor that can be fixed to the white cane. Ultrasonic ETAs such as the "K"-Sonar (BAT Technologies) and the Miniguide (GDP Research) are still being manufactured (Smith and Penrod, 2010), notably the UltraCane (Sound Foresight Technology), which combines two ultrasonic sensors to a traditional white cane (Hoyle and Waters, 2008), and the newer WeWALK smart cane (WeWALK, 2019), an innovative "all-in-one" primary aid. It combines the traditional white cane with a single ultrasonic sensor, a touch pad and a voice assistant for smart control of the user's smartphone without requiring the other hand. Examples of ETAs are depicted in **Figure 4**.

Modern Sensory Substitution

Non-invasive sensory substitution endeavors to use a non-visual sensory input to stimulate the visual cortex and other brain areas related to vision, all *via* natural rerouting of existing

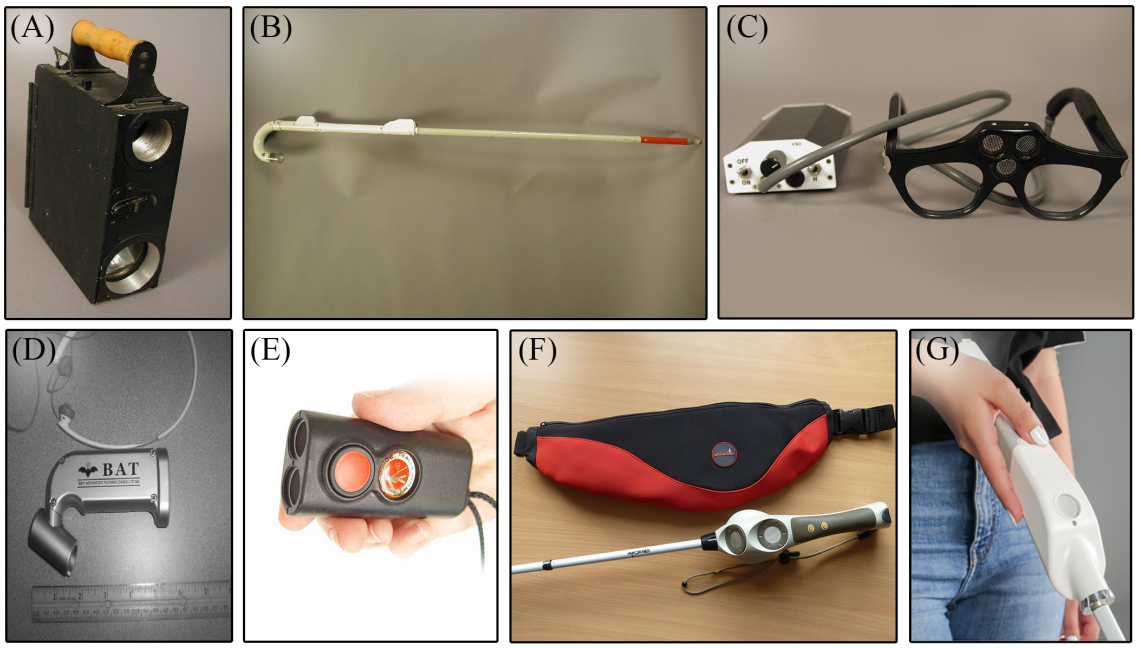

FIGURE 4 | Non-invasive aids. On the upper row are historic ETAs that are no longer available (photo courtesy AER O&M Division Warren Bledsoe Archives, American Printing House for the Blind). **(A)** Signal Corps Obstacle Detector; **(B)** C-4 laser cane; **(C)** Sonic Guide; **(D)** "K"-Sonar [from Penrod et al. (2009); with the permission of W. Penrod]; **(E)** Miniguide (GDP Research); **(F)** UltraCane (from http://www.ultracane.com, with the permission of Sound Foresight Technology Ltd); and **(G)** WeWALK smart cane (from https://wewalk.io/en/, with the permission of WeWALK Tech Co.).

sensory channels (Foulke et al., 1986). These methods thus exploit the brain's natural adaptation mechanisms. They offer new possibilities to "restore" visual function in blind people, and have attracted considerable interest since their inception. Paul Bach-y-Rita did the pioneering work on sensory substitution in the 1970s. At a time when most scientists believed the visual areas of the blind to be atrophied and non-functional, Bach-y-Rita argued that the visually deprived brain could readapt, since it had only lost the peripheral systems (i.e., eye, retina). In *Brain Mechanisms in Sensory Substitution*, Bach-y-Rita (1972) recounted that the images captured by the eyes travel to the brain in the form of neuronal signals. Therefore, sight is not mediated by the eyes, but by the brain's interpretation of incident electrical signals, based on hard-wiring of the brain, but also informed by memory, learning, contextual interpretations, and many other factors (Bach-y-Rita and Aiello, 1996). According to Bach-y-Rita, people living with blindness could regain access to the missing visual input if only were made accessible *via* their intact senses (Bach-y-Rita et al., 1969). With this idea in mind, Bach-y-Rita designed the Tactile Vision Substitution System (TVSS), a sensory substitution system for transmitting visual information through the skin surface of the back. A camera captured visual information that was then transmitted over an electro-tactile grid which activated skin receptors that sent visual information to the brain, where it is processed and perceived. Case study investigations showed us that it is possible, with some learning, to feel and interpret different patterns drawn on the skin of the back and then to use that information to judge distances and even catch moving objects. Several models of the TVSS were manufactured with the goal of greater portability and increased effectiveness in the visual domain. Bach-y-Rita investigated the use of electrodes arrays on the fingers (Kaczmarek et al., 1994), on the abdomen (Kaczmarek et al., 1985) and on the tongue (Bach-y-Rita et al., 1998; Sampaio et al., 2001). He concluded that the tongue was the best option based on several criteria. First, the tactile sensitivity of the tongue is significantly greater than that of the skin of the back or fingers. Second, the cortical surface for the tongue is larger than the corresponding surface for the entire back. Third, the tongue's tactile receptors are closer to its surface, while the saliva, which is an electrolytic solution, assures electrical contact between the electrodes and the tongue (Bach-y-Rita, 2004). Consequently, the tongue requires significantly less voltage and current than does the fingertip in order to perceive electrotactile stimulations (Bach-y-Rita et al., 1998; Bach-y-Rita and Kaczmarek, 2002).

First Generation of SSDs: TDU, vOICe, and PSVA

The Tongue Display Unit (or TDU) transmits visual input to the tongue in the form of electrotactile pulses. It is composed of a 20 × 20 matrix array of small circular electrodes that is placed on the tongue, a laptop computer and a webcam attached to eyeglasses. The visual image is translated into electrotactile pulses and thus "drawn" in real time with the application of electrical currents on the tongue. Several studies have shown that TDU allows the blind to perceive light sources (Nau et al., 2013; Lee et al., 2014), movement (Ptito et al., 2009; Matteau et al., 2010), and to recognize shapes (Ptito and Kupers, 2005; Vincent et al., 2014), objects (Williams et al., 2011; Nau et al., 2015b), and letters

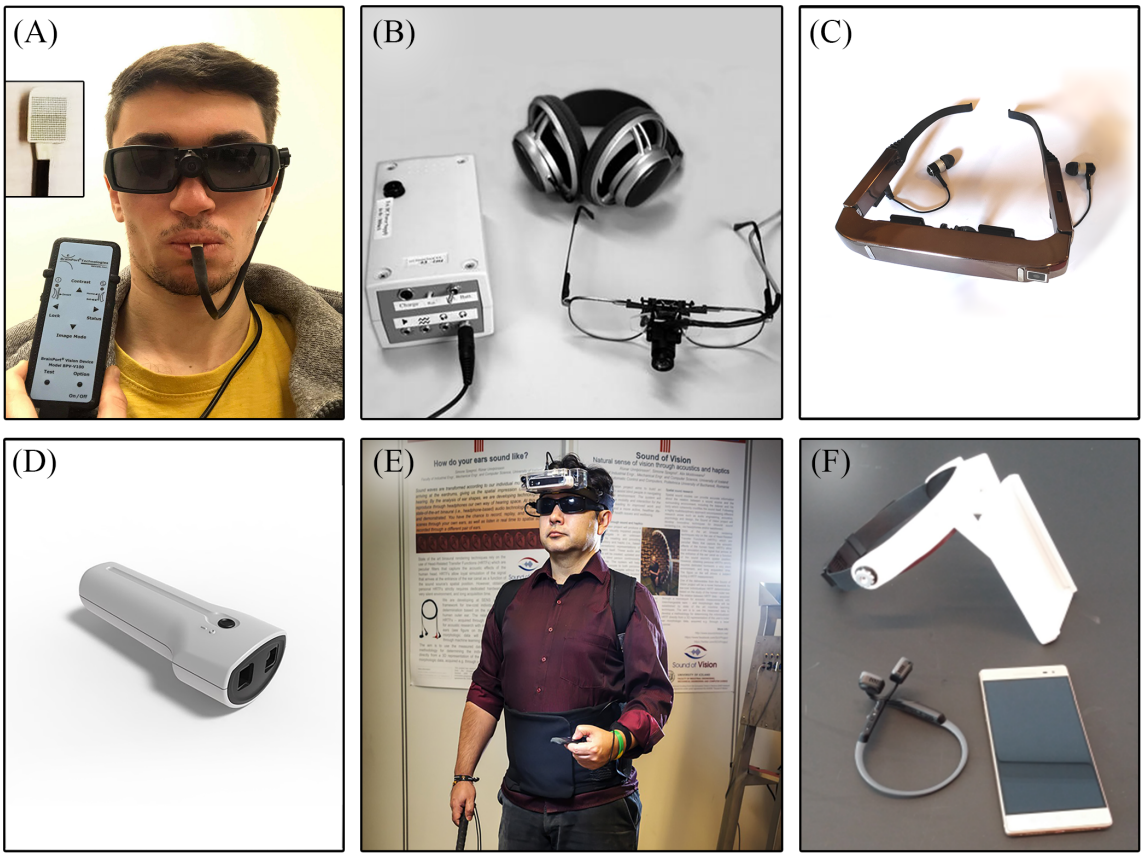

FIGURE 5 | Illustrations of SSDs. **(A)** Tongue Display Unit (TDU); **(B)** vOICe; **(C)** Prosthesis for Substitution of Vision with Audition (PSVA) [from Collignon et al. (2007); with the permission of O. Collignon]; **(D)** Eyecane; **(E)** Sound of Vision (SoV) [adapted from Hoffmann et al. (2018)]; and **(F)** Guidance Sensory Substitution Device (GSSD).

(Chebat et al., 2007b; Pamir et al., 2020). Users of the TDU are even able to navigate in an obstacle course (Chebat et al., 2011). Furthermore, the estimated "visual acuity" of the tongue attained an acuity of 1/90 in trained users (Chebat et al., 2007b), which meets the criterion of low vision that is sufficient to perceive environmental shapes (Ptito et al., 2012) and useful for many visual tasks [reviewed in Stronks et al. (2016)].

Since blind people are also able to perform certain spatial tasks using sound cues (Kellogg, 1962; Bassett and Eastmond, 1964), auditory SSDs have been developed to enhance this skill (Meijer, 1992; Capelle et al., 1998; Bronkhorst and Houtgast, 1999; Kay, 2000). The best known auditory-to-vision SSD was described by Meijer (1992), who named their device the vOICe, where the capitals O, I, and C represent the exclamation "Oh, I see!" The system offers "functional vision" by converting images captured by a video camera to different soundscapes. To do so, the algorithm uses a scanning technique that divides the field of view (FOV) into a matrix of pixels. Initially, the system used a 64 by 64 pixels matrix containing 4,096 elements, but has since evolved to generate a much higher resolution of up to 25,344 pixels (Striem-Amit et al., 2012b). The algorithm analyses every column of pixels in a left to right sequence to translate vertical position to the frequency domain and horizontal position to the duration of the sound. As for colors, they are integrated in a scale of 16 shades of gray so the system can convert luminosity to different sound amplitudes (Meijer, 1992). Striem-Amit et al. (2012b) evaluated the audio-visual acuity of the vOICe users after receiving 73 h of training with the device, more than half of whom had attained a visual acuity of 20/320 which outclasses the threshold of blindness (20/400) defined by the World Health organization. Moreover, several studies with the vOICe demonstrated that blind individuals can learn to identify geometric forms and shapes (Amedi et al., 2007), read (Striem-Amit et al., 2012a), locate objects in space (Auvray et al., 2007) and even learn virtual maps (Jicol et al., 2020). Since then, a new version of the vOICe has been developed to add color information to the mixture of the visual information given by the device. Named the Eyemusic, it performs a spectral analysis of the image and links specific colors with recordings of different musical instruments (Abboud et al., 2013, 2014). Therefore, the device simultaneously conveys spatial information and color, thus enhancing the user's comprehension of space. Furthermore, blind individuals trained with the device were able to recognize facial expressions with the device (Abboud et al., 2013; Arbel-Yaffe and Amedi, 2016).

Another auditory-to-vision SSD known as the prosthesis for substitution of vision with audition (PSVA), has a field of view

(FOV) divided in a differential resolution structure, in which the center contains additional pixels for a higher resolution thus mimicking the human retina and its fovea, which mainly serves for pattern recognition while lower resolution in the periphery allows spatial localization and movement detection (Capelle et al., 1998). The PSVA offers a sonification strategy similar to that of the vOICe by assigning each pixel a sinusoidal tone at distinct pitches and modulated by the gray level intensity. However, instead of scanning images, it uses binaural differences and tone intensity to code for horizontal positioning, while different frequencies are used for vertical positioning, thus, exploiting the natural mechanisms of human hearing (Gulick et al., 1989). Few behavioral studies have been done with this device. However, studies in blind individuals have shown that the PSVA imparts efficient pattern recognition (Arno et al., 2001), spatial localization (Collignon et al., 2007), and depth perception (Renier et al., 2005b).

New Generation of SSDs: Eyecane, SoV, and GSSD

The newer generation of SSDs is not designed to restore high resolution vision, but rather to gather and transmit specifically chosen cues to provide greater independence to the user in a specific task such as navigation. The Eyecane, for example, is a minimalist SSD that uses a "point-to-distance" technology as an aid to navigate. In brief, the device uses infrared light sensors to detect a single point in front of the user and calculate the distance between the detected obstacle and its sensor. The device then conveys this information in the form of tactile (vibrations) and auditory cues such as, higher the vibrations and sounds, as one approaches the object (Maidenbaum et al., 2014b). With its small and handy structure, it is designed to bring greater freedom than is afforded by the white cane while also providing superior detection range (Buchs et al., 2014, 2017; Maidenbaum et al., 2014c). This device enables quick and efficient perception of the distance between the user and obstacles in the environment by using sweeping motions, analogous to those with the white cane, thus requiring minimal additional training. Using this device, CB participants were able to navigate in a Hebb–Williams maze as efficiently as sighted participants (Chebat et al., 2015), and were able to transfer spatial information from a virtual environment to the real world (Chebat et al., 2017).

Another promising navigational aid called the Sound of Vision (or SoV) was recently developed. The SoV uses a combination of sensors and a video camera (both mounted on the forehead) to convey the 3D information of the environment, namely depth, positioning, form, and size, *via* a hybrid audio-haptic signal. The haptic signal is delivered on the skin of the abdomen to inform the user of the closest obstacle (Caraiman et al., 2017). As for the auditory signal, the system divides its FOV into a 3 by 5 matrix, in which every sector of the matrix codes and translates depth and direction information into spatialized "popping bubbles" sounds. Thereby, the user can extract the form and the position of an obstacle, while estimating its distance (Hoffmann et al., 2018). Interestingly, the SoV system simplifies its signal by encoding only the closest obstacles in the user's path thus reducing the cognitive demand placed on the user (Caraiman et al., 2017).

Since smartphones are increasing in popularity in the blind community (Kacorri et al., 2017), SSDs have come to exploit their accessibility and simplicity by making available useful applications. Once such novel smartphone application called the Guidance-Sensory-Substitution-Device (or GSSD) guides users through obstacles, thus increasing their navigational independence. The GSSD uses the cameras of smartphones to capture the environment and bone-conducting earphones to inform the individual of oncoming obstacles by broadcasting horizontally spatialized sounds. The GSSD conveys a simple auditory output based on the point-to-distance principle, while signaling every potential obstacle with a singular sound source that depicts the distance of closest edges from the user. By this means, the user can associate each sound source to a specific obstacle and then plan her/his route through space (Paré et al., 2019). Illustrations of the SSDs are shown in **Figure 5**.

SENSORY SUBSTITUTION AND CROSS-MODAL REWIRING OF THE BRAIN IN CONGENITAL AND LATE BLINDNESS

Sensory Substitution

Studies on sensory substitution in CB concur in showing their superior spatio-cognitive skills, which again show that the blind have come to possess certain supernormal skills for sound localization (Lessard et al., 1998) and proprioception (Loomis et al., 1993). In addition to spatial tasks, several other studies also show a marked perceptual advantage for performing cognitive tasks (Muchnik et al., 1991; Röder et al., 1999; Bavelier and Neville, 2002; see also Kupers and Ptito, 2014), verbal memory (Amedi et al., 2003), and attention (Muchnik et al., 1991; Röder et al., 1996, 1999; Liotti et al., 1998). Since the pioneering work demonstrating that the visual cortex of CB can, with training, be recruited by tactile stimulation, i.e., training-induced plasticity (Sadato et al., 1996; Ptito et al., 2005), the bulk of subsequent studies has confirmed the activation of the visual cortex in tactile, auditory, and olfactory tasks [reviewed in Kupers and Ptito (2014) and Nau et al. (2015a)]. Interestingly enough, not only is the visual cortex activated by tactile stimuli but the tactile motion and shape information are funneled into the dorsal (Ptito et al., 2009) and ventral visual pathways (Ptito et al., 2012), respectively. This phenomenon has also been shown upon auditory stimulation of encoded visual information (Collignon et al., 2007; Striem-Amit and Amedi, 2014; Arbel-Yaffe and Amedi, 2016). This recruitment of visual areas for tactile and auditory tasks gives CBs an advantage for the use of sensory substitution devices (Ptito et al., 2005), and allows them to significantly increase their performance after only a few hours of training (Sampaio et al., 2001). Moreover, the brain areas activated when exploring a virtual maze using a tactile-to-vision substitution device roughly matched the areas activated when sighted people explored a virtual maze using vision, but differed from those activated in blindfolded sighted controls. We have previously shown that the occipital cortex

and the hippocampal/parahippocampal complex are involved in route recognition in CBs, similar to sighted people performing the same tasks with opened eyes (Kupers et al., 2010; Chebat et al., 2020). This network of brain regions is important for navigational behavior in sighted people (Maguire et al., 1998, 2000; Schindler et al., 2004; Epstein, 2008; Browning et al., 2009; Squire, 2009). These natural mechanisms of adaptation in the blind brain should be used to guide the development of training programs using SSDs, since they highlight the inherent ability of the brain to recruit task-specific areas when using substituted sense modalities (Chebat et al., 2018b).

Cross-Modal Plasticity

Since congenital blindness and early onset of vision loss alters the retinofugal projections to the visual cortex, the blind brain undergoes a massive anatomical reorganization leading to cross-modal plastic reconfigurations of sensory pathways. This is possible because the brain has a natural ability, called neuroplasticity, to adapt itself in response to every perturbation in the external and the internal environment. The first structural/functional studies on the visual system of blind people using magnetic resonance imaging (MRI) and positron emission tomography (PET) found significant alterations not only in the white matter tracts including the optic nerves, the optic chiasm and the optic tracts (Breitenseher et al., 1998; Ptito et al., 2008) but also relative reductions of the gray matter volume in the visual thalamus (the lateral geniculate nucleus, and posterior pulvinar), and striate and extra-striate visual cortices (Shimony et al., 2006; Ptito et al., 2008; Cecchetti et al., 2016). Other volumetric reductions were reported in the brain commissural systems such as the splenium of corpus callosum (Ptito et al., 2008; Tomaiuolo et al., 2014; Cavaliere et al., 2020), accompanied by an enlargement of the anterior commissure (Cavaliere et al., 2020). In addition, regions connected to the dorsal visual stream such as the hippocampus were also reduced in volume (Chebat et al., 2007a; Fortin et al., 2008). Cortical thickness is increased in the primary visual cortex of the congenitally blind (Jiang et al., 2009; Kupers et al., 2011) accompanied by a supra-metabolic activity therein (De Volder et al., 1997; Kupers and Ptito, 2014). **Figure 6** shows the atrophy in various components of the visual system of CB individuals.

Furthermore, magnetoencephalography has provided evidence for increased functional connectivity of the occipital cortex with auditory and somatosensory areas (Ioannides et al., 2013; Kupers and Ptito, 2014; Müller et al., 2019), as likewise shown in studies using transcranial magnetic stimulation (Wittenberg et al., 2004; Kupers et al., 2006). Other functional connectivity studies revealed stronger connections of the visual cortex with somatosensory (Shu et al., 2009), auditory (Watkins et al., 2013; Burton et al., 2014), and language areas (Bedny et al., 2011; Butt et al., 2013). Finally, a recent resting state functional magnetic resonance imaging (rsfMRI) study (Heine et al., 2015) revealed increased functional connectivity within both the ventral and the dorsal visual streams in congenitally blind participants along with a stronger functional connectivity between the occipital cortex and language areas, and regions involved in tactile (Braille) processing such as the inferior frontal gyrus (Broca's area), the thalamus, the supramarginal gyrus and the cerebellum (Heine et al., 2015). Taken together, most anatomical studies concur in showing that the tactile or auditory information reach the visual cortex of the blind through both a multi-synaptic cortico-cortical pathway (Ptito and Kupers, 2005) and also through a direct thalamo-cortical pathway (Kupers and Ptito, 2014; Murphy et al., 2016; Müller et al., 2019). The cross-modal rewiring of the blind brain is illustrated in **Figure 7**.

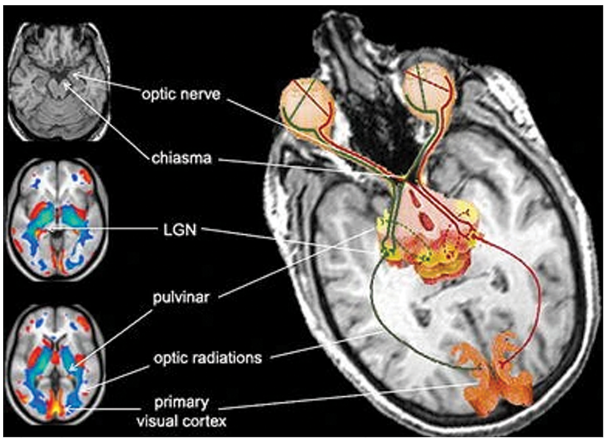

FIGURE 6 | Atrophy of the components of the visual system in congenitally blind individuals [from Ptito et al. (2008); with the permission of Springer Nature, license # 4986491335589].

Late Blindness

The study of late acquired blindness (LB) poses a completely different challenge than early acquired blindness (Chebat et al., 2018b). LB subjects have a visual system that has developed normally until vision loss and basically, they possess a visual brain similar to that of seeing people. Two important parameters were and still are often neglected in studies on late blindness, namely the onset and duration of blindness, which led to the contradictory results reported in the literature. As of now, most of the studies on sensory substitution only tested CB individuals or a mix of LB subjects without considering onset and duration of blindness. It is known that the neuroplastic processes that accompany the onset of blindness are less strong in LB, taking into account that plasticity is highly dependent on critical periods of development (Sadato et al., 2002; Noppeney, 2007; Jiang et al., 2009). One could therefore argue that once this critical period is over, the brain is less likely to adapt itself to a new condition. Nonetheless, a number of studies have reported neuroanatomical differences between CB and LB, and LB and subjects with normal vision, which challenges the rigidity of critical periods in the brain (Heimler and Amedi, 2020). For example, cross-modal plastic processes have been usually found in CB whose visual cortex is activated by other senses like audition, touch and even smell [reviewed in Kupers and Ptito (2014)]. These plastic manifestations are also found in LB but in the extra-striate visual areas (Sadato et al., 2004; Renier et al., 2005a; Amedi et al., 2007;

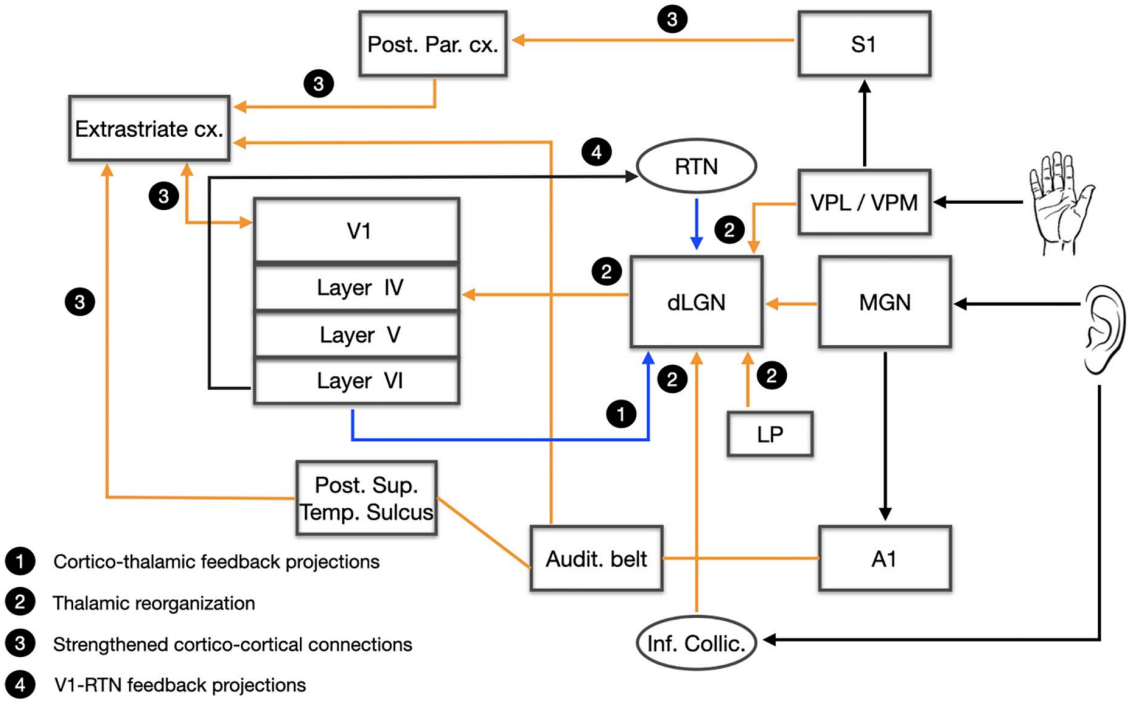

FIGURE 7 | The rewired blind brain. A schematic representation of the reorganization of the blind brain. Four networks are presented: **(1)** Cortico-thalamic feedback projections; **(2)** Thalamic reorganization; **(3)** Strengthened cortico-cortical connections; **(4)** V1-RTN feedback projections. Audit. Belt, auditory belt; A1, primary auditory cortex; Cx, cortex; dLGN, dorsal lateral geniculate nucleus; LP, lateral pulvinar; MGN, medial geniculate nucleus; Post. Par. Cx.; posterior parietal cortex; Post. Sup. Temp. Sulcus, posterior superior temporal sulcus; RTN, reticular thalamic nucleus; S1, primary somatosensory cortex; VPL, ventral posterolateral nucleus; VPM, ventral posteromedian nucleus; V1, primary visual cortex.

Collignon et al., 2013) and in the splenium of the corpus callosum (Shi et al., 2015; Cavaliere et al., 2020).

Moreover, only a handful of studies have been devoted to the perceptual, cognitive and navigational abilities of late blind individuals (Chebat et al., 2018b). Differences were shown mainly in auditory capacities and navigational strategies when compared to CB [reviewed in Kupers and Ptito (2014)]. For instance, LB have inferior abilities than CB in using binaural and monaural cues for localizing sound sources (Voss et al., 2004) and in echolocation (Dufour et al., 2005) but have better performances in auditory spatial bisection (Amadeo et al., 2019). Moreover, before vision loss, subjects learn to navigate using mostly allocentric strategies. Without vision, LB has to adapt their strategies by transiting into egocentric point of views with only tactile and auditory cues like CB individuals do. Although LB can learn to use SSDs very efficiently (Lee et al., 2014; Chebat et al., 2015, 2017; Paré et al., 2019), it is clear that they do not possess the same skills as CB (Wan et al., 2010; Chebat et al., 2015, 2017). This is probably due to the fact that the cross-modal changes witnessed in the late blind are limited compared to that of CB (Park et al., 2009; Reislev et al., 2017; Wen et al., 2018). Therefore, the visual experience of LB seems to impair their ability to use SSDs compared to CB and their visual experience seems to be detrimental to cross-modal rewiring of the brain. Invasive devices however, are geared specifically toward LB since their technology requires visual experience (Castaldi et al., 2016).

DISCUSSION

Brain-Machine Interfaces to Assist the Blind

In this chapter, we briefly described the history of blindness from ancient to modern times. We then addressed the various means that have been used to help blind individuals throughout history, with an emphasis on modern technologies. We divided these aids into two categories: invasive prostheses and non-invasive brain interfaces.

Invasive Techniques and Their Limitations

The retina and the visual cortex have been the site of choice for most of the visual prostheses employing electrical stimulation. Located at both extremes of the visual pathways, they are more surgically accessible than are deep brain structures such as the optic nerve and the LGN. Targeting these terminal sites presents certain advantages and challenges. In general, the electrical stimulation of the visual pathways induces phosphenes. In epiretinal prosthesis, the evoked phosphenes have proven to be highly variable and dependent on the activation of passing axon fibers by the implanted electrodes (Beyeler et al., 2019). Moreover, the retina undertakes complex processing of visual inputs, extending from the spatiotemporal integration of light by the photoreceptors to the output of RGCs to the deep visual relay centers (Demb and Singer, 2015). Therefore,

stimulation strategies should take into account the structural and functional properties of the retina in order to reproduce a naturalistic activity in the RGC layer for downstream processing in cortical visual areas (Nirenberg and Pandarinath, 2012). The visual cortex is the primary recipient of the retino-geniculate input, which is then processed further in higher order visual areas. However, the neuronal and processing complexity is much higher therein, making it difficult to obtain a meaningful perception through electrical stimulation of the retina only. A major limitation of this approach is that retinal neurons activation affects the activation/inhibition balance that influences the signal propagation to higher order cortical areas (Bosking et al., 2017). While the visual cortex was the first site of stimulation to be explored, it took longer time to reach the safety standards required for human clinical trials, given the obviously more invasive surgical procedures involved. This is why most of the clinical trials have hitherto employed retinal prostheses that lead to letter and object recognition and navigation. Moreover, the best visual acuity offered to date by visual prostheses still falls below the threshold of visual acuity that defines blindness (20/400). For the present, the surgical risks remain too great to justify the few benefits provided by invasive prostheses. Indeed, major neurosurgical procedures are inherently dangerous and can cause deleterious complications such as infection, inflammation, and neurodegeneration along with other neurological problems. Another element restricting the use of these technologies is that they are not appropriate for people who were deprived of vision since birth. Their efficacy relies on the presence of a normally developed visual system with a visual repertoire acquired through experience (Reich et al., 2012). In CB, who were deprived of visual inputs since birth, the visual system undergoes cross-modal rewiring that leads to a massive reorganization of non-visual inputs to the visual cortex (see **Figure 8**) [reviewed in Kupers and Ptito (2014)] which disfavors the use of surgical prostheses.

Although electrical stimulation has been extensively used in experimental setting, other stimulation strategies are under investigation. The new technique of optogenetics uses viral vectors to genetically modify cells to express rhodopsin, enabling the modulation of neuronal population activity by light with high spatiotemporal resolution. This technique has been explored both for the retina and the visual cortex. Current clinical trials are testing the feasibility of using optogenetics to render the RP patients sensitive to light (Farnum and Pelled, 2020).

Non-invasive Devices

Sensory substitution and electronic aids have an advantage over invasive technologies by virtue of exploiting the plasticity mechanisms that naturally operate in the blind brain when trained in other modalities. While some SSDs provide the blind a "visual" perception that exceeds the World Health Organization legal blindness threshold and with no health risks, several factors limit their use outside laboratories. For instance, the spatial resolutions of available devices are limited by the targeted sensory modality. Indeed, since hearing and touch both have lesser spatial bandwidth than natural vision (Bach-y-Rita et al., 1969; Bach-y-Rita, 1972; Apkarian-Stielau and Loomis, 1975; Wiley et al., 1986; Ashmead et al., 1990), a direct translation of visual information to either touch or hearing inevitably results in loss of details (Loomis et al., 2012). Moreover, SSDs are generally designed to assist the blind without consideration of their opinions, contribution and cooperation, and have only been validated in heterogeneous

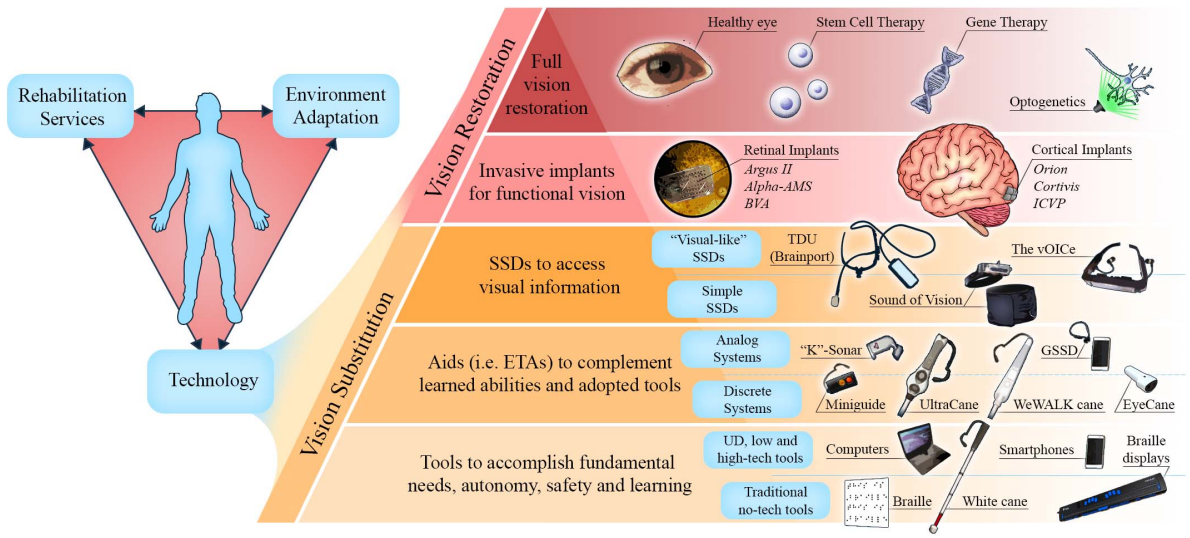

FIGURE 8 | Technologies to assist the blind. (Left) Blind individuals can gain autonomy *via* a combination of the following fundamental and interlinked factors: rehabilitation services, environment adaptation, and technology. (Right) A model of present and future technologies and therapies to substitute and restore vision. At the bottom of the pyramid are simple devices, or tools, that are currently adopted by blind individuals to meet their fundamental needs. At each level, technologies aim to enable more tasks toward the goal of full vision substitution or full vision restoration. However, due to limitations discussed in this review, the higher the technology or therapy on the pyramid, the greater the obstacles to its application and adoption by blind individuals. ETAs, electronic travel aids; SSDs, sensory substitution devices; UD, universal design.

populations of late onset and congenitally blind individuals. This obviously impacts the results on behavior [reviewed in Kupers and Ptito (2014) and Chebat et al. (2018a)]. Important factors that could influence the way SSDs are used and appreciated by users have hitherto been underestimated (Elli et al., 2014). As a consequence, numerous devices have proven to be either too complex or too expensive to operate in real life situations. Indeed, many devices require several hours if not days and months of training that discourage the blind for using them (i.e., the vOICe). This is a major impediment to their broader implementation since most of the attentional resources of the users are focused on decoding the SSD signal instead of understanding their surroundings. This attentional misplacement leads to cognitive overload and exhaustion in complex environments (Elli et al., 2014; Pissaloux and Velázquez, 2018). Indeed, Consequently, the blind community in general is not highly motivated to adopt these apparatuses (Elli et al., 2014; Maidenbaum et al., 2014a; Chebat et al., 2018a).

A more compelling solution for individuals living with blindness is presented by the new minimalist SSDs (Eyecane, GSSD) and ETAs (Miniguide, UltraCane, and WeWALK cane), which are the mainstays of assistive mobility technologies currently used and introduced in O&M training (Smith and Penrod, 2010). Their broader application is favored by the greater simplicity of their signals and ease of use, which makes them acceptable supplementation aids. Furthermore, the advent of computers and smartphones with accessible software (built with universal design) allows more flexibility and opportunities for individuals to share their experiences with the rest of the population. As an example, screen-reading software, optical character recognition software, and travel related applications adapted for the blind can all be accessed through smartphones, and have become increasingly popular amongst individuals with blindness (Kacorri et al., 2017). As in the case of the GSSD (Paré et al., 2019), sensory substitution could also benefit from the processing capacities of smartphones by being designed as downloadable smartphone applications. There is also scope for adapting the urban environment better to suit the needs of individuals living with disabilities, and to increase their safety and autonomy as stated in *American Disability Act* (ADA, 1990). Indeed, the increasing number of measures such as the installation of tactile plates and auditory pedestrian signals are good examples of such universal design. This calls for the promotion of widespread standardization of such enabling measures, and also calls for further research and development of technologies, like presenting 3D printed tactile maps in buildings and in public places. Moreover, artificial intelligence is a promising venue as it can provide blind individuals with devices or applications equipped with image recognition software for text, faces, objects, and even larger scale environments to enable more efficient interactions and autonomous mobility (Morrison et al., 2017; Kelley, 2018; Zhao et al., 2018). Our view is illustrated in **Figure 8**, which highlights the present and future methodologies extending from simple vision substitution to full vision restoration through highly sophisticated interventions such as gene therapy, stem cell technology or optogenetics.

ETHICS STATEMENT

Written, informed consent was obtained from the participants for the publication of any identifiable data and images.

AUTHOR CONTRIBUTIONS

MP, D-RC, and FS planned the review. MP, D-RC, FS, SP, MB, and ID wrote the manuscript. All authors have contributed equally to this work.

FUNDING

This work was supported by the Harland Sanders Research Chair in Visual Science (University of Montreal; MP, MB, SP, and ID), Réseau de recherche en Santé de la Vision du Québec (MB), the Ariel University Research Authority Absorption Grant # RA1700000192 (D-RC), and the French Health Ministry PHRC 0701044 (FS).

REFERENCES

Abboud, S., Hanassy, S., Levy-Tzedek, S., Maidenbaum, S., and Amedi, A. (2014). EyeMusic: introducing a "visual" colorful experience for the blind using auditory sensory substitution. *Restor. Neurol. Neurosci.* 32, 247–257. doi: 10.3233/RNN-130338

Abboud, S., Hanassy, S., Levy-Tzedek, S., Tauber, D., Maidenbaum, S., and Amedi, A. (2013). EyeMusic: a colorful experience for the blind. *Multisens. Res.* 26:116. doi: 10.1163/22134808-000s0084

Ahuja, A. K., Dorn, J. D., Caspi, A., McMahon, M. J., Dagnelie, G., daCruz, L., et al. (2011). Blind subjects implanted with the Argus II retinal prosthesis are able to improve performance in a spatial-motor task. *Br. J. Ophthalmol.* 95, 539–543. doi: 10.1136/bjo.2010.179622

Allen, P. J., and Ayton, L. N. (2020). "Development and Experimental Basis for the Future of Prosthetic Vision," in *Macular Surgery* eds A. Chang, M. Ohji and W. F. Mieler (Berlin: Springer), 449–462. doi: 10.1007/978-981-15-7644-7_33

Amadeo, M. B., Campus, C., and Gori, M. (2019). Impact of years of blindness on neural circuits underlying auditory spatial representation. *Neuroimage* 191, 140–149. doi: 10.1016/j.neuroimage.2019.01.073

Amedi, A., Raz, N., Pianka, P., Malach, R., and Zohary, E. (2003). Early 'visual'cortex activation correlates with superior verbal memory performance in the blind. *Nat. neurosci.* 6, 758–766. doi: 10.1038/nn1072

Amedi, A., Stern, W. M., Camprodon, J. A., Bermpohl, F., Merabet, L., Rotman, S., et al. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nat. Neurosci.* 10, 687–689. doi: 10.1038/nn1912

Apkarian-Stielau, P., and Loomis, J. M. (1975). A comparison of tactile and blurred visual form perception. *Percept. Psychophys.* 18, 362–368. doi: 10.3758/bf03211213

Arbel-Yaffe, R., and Amedi, A. (2016). Auditory face identification activates selective areas within the ventral visual stream in congenitally blind. *J. Vis.* 16:1229. doi: 10.1167/16.12.1229

Arno, P., De Volder, A. G., Vanlierde, A., Wanet-Defalque, M. C., Streel, E., Robert, A., et al. (2001). Occipital activation by pattern recognition in the

early blind using auditory substitution for vision. *Neuroimage* 13, 632–645. doi: 10.1006/nimg.2000.0731

Arsiero, M., Cruz, L. D., Merlini, F., Sahel, J. A., Stanga, P. E., Hafezi, F., et al. (2011). Subjects blinded by outer retinal dystrophies are able to recognize shapes using the argus II retinal prosthesis system. *Invest. Ophthalmol. Vis. Sci.* 52, 4951–4951.

Ashmead, D. H., Leroy, D., and Odom, R. D. (1990). Perception of the relative distances of nearby sound sources. *Percept. Psychophys.* 47, 326–331. doi: 10.3758/bf03210871

Auvray, M., Hanneton, S., and O'Regan, J. K. (2007). Learning to perceive with a visuo—auditory substitution system: localisation and object recognition with 'the voice'. *Perception* 36, 416–430. doi: 10.1068/p5631

Ayton, L. N., Blamey, P. J., Guymer, R. H., Luu, C. D., Nayagam, D. A. X., Sinclair, N. C., et al. (2014). First-in-human trial of a novel suprachoroidal retinal prosthesis. *PLoS One* 9:e115239. doi: 10.1371/journal.pone.0115239

Bach-y-Rita, P. (1972). *Brain Mechanisms in Sensory Substitution*. Cambridge: Academic Press Inc.

Bach-y-Rita, P. (2004). Tactile sensory substitution studies. *Ann. N. Y. Acad. Sci.* 1013, 83–91. doi: 10.1196/annals.1305.006

Bach-y-Rita, P., and Aiello, G.L. (1996). Nerve length and volume in synaptic vs diffusion neurotransmission: a model. *NeuroReport* 7, 1502–1504. doi: 10.1097/00001756-199606170-00012

Bach-y-Rita, P., Collins, C. C., Saunders, F. A., White, B., and Scadden, L. (1969). Vision substitution by tactile image projection. *Nature* 221, 963–964. doi: 10.1038/221963a0

Bach-y-Rita, P., and Kaczmarek, K. A. (2002). Tongue placed tactile output device. *US Patent* 6, 430–459.

Bach-y-Rita, P., Kaczmarek, K. A., Tyler, M. E., and Garcia-Lara, J. (1998). Form perception with a 49-point electrotactile stimulus array on the tongue: a technical note. *J. Rehabil. Res. Dev.* 35, 427–430.

Barasch, M. (2001). *Blindness: the History of a Mental Image in Western Thought*. Abingdon: Routledge.

Barry, M. P., Armenta Salas, M., Patel, U., Wuyyuru, V., Niketeghad, S., Bosking, W. H., et al. (2020). Video-mode percepts are smaller than sums of single-electrode phosphenes with the Orion® visual cortical prosthesis. *Invest. Ophthalmol. Vis. Sci.* 61:927.

Bassett, I. G., and Eastmond, E. J. (1964). Echolocation: measurement of pitch versus distance for sounds reflected from a flat surface. *J. Acoust. Soc. Am.* 36:911. doi: 10.1121/1.1919117

Bavelier, D., and Neville, H. J. (2002). Cross-modal plasticity: where and how? *Nat. Rev. Neurosci.* 3, 443–452. doi: 10.1038/nrn848

Bedny, M., Pascual-Leone, A., Dodell-Feder, D., Fedorenko, E., and Saxe, R. (2011). Language processing in the occipital cortex of congenitally blind adults. *Proc. Natl. Acad. Sci.* 108, 4429–4434. doi: 10.1073/pnas.1014818108

Beggs, W. (1992). Coping, adjustment, and mobility-related feelings of newly visually impaired young adults. *J. Vis. Impair. Blind.* 86, 136–140. doi: 10.1177/0145482x9208600307

Benjamin, J. (1968). A review of the veterans administration blind guidance device project. *Bull. Prosthet. Res.* 10:64.

Benjamin, J. M. (1974). The laser cane. *Bull. Prosthet. Res.* 22–24, 443–450.

Beyeler, M., Nanduri, D., Weiland, J. D., Rokem, A., Boynton, G. M., and Fine, I. (2019). A model of ganglion axon pathways accounts for percepts elicited by retinal implants. *Sci. Rep.* 9:9199. doi: 10.1038/s41598-019-45416-4

Bledsoe, C. (2010). "The originators of orientation and mobility training,". in *Foundations of Orientation and Mobility* eds L. W. Richard, B. B. Bruce and R. W. William. New York: AFB Press).

Bliss, J. C., King, B., Kotovsky, K., and Crane, H. D. (1963). *Tactual Perception of Visual Information*. Menlo Park Calif: Stanford Research Inst).

Boehm, R. (1986). The use of echolocation as a mobility aid for blind persons. *J. Vis. Impair. Blind.* 80, 953–954.

Bosking, W. H., Beauchamp, M. S., and Yoshor, D. (2017). Electrical stimulation of visual cortex: relevance for the development of visual cortical prosthetics. *Ann. Rev. Vis. Sci.* 3, 141–166. doi: 10.1146/annurev-vision-111815-114525

Breitenseher, M., Uhl, F., Wimberger, D. P., Deecke, L., Trattnig, S., and Kramer, J. (1998). Morphological dissociation between visual pathways and cortex: MRI of visually-deprived patients with congenital peripheral blindness. *Neuroradiology* 40, 424–427. doi: 10.1007/s002340050616

Brindley, G. S., and Lewin, W. S. (1968). The sensations produced by electrical stimulation of the visual cortex. *J. Physiol.* 196, 479–493. doi: 10.1113/jphysiol.1968.sp008519

Bronkhorst, A. W., and Houtgast, T. (1999). Auditory distance perception in rooms. *Nature* 397, 517–520. doi: 10.1038/17374

Browning, N. A., Grossberg, S., and Mingolla, E. (2009). Cortical dynamics of navigation and steering in natural scenes: motion-based object segmentation, heading, and obstacle avoidance. *Neural Netw.* 22, 1383–1398. doi: 10.1016/j.neunet.2009.05.007

Buchs, G., Maidenbaum, S., and Amedi, A. (2014). "Obstacle identification and avoidance using the 'EyeCane': a tactile sensory substitution device for blind individuals", in *Proceedings of the International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*: Berlin: Springer), 96–103. doi: 10.1007/978-3-662-44196-1_13

Buchs, G., Simon, N., Maidenbaum, S., and Amedi, A. (2017). Waist-up protection for blind individuals using the EyeCane as a primary and secondary mobility aid. *Restor. Neurol. Neurosci.* 35, 225–235. doi: 10.3233/rnn-160686

Burklen, K. (1924). *Blinden Psychologie*. Leipzig: verlag von johann ambrosius barth.

Burton, H., Snyder, A. Z., and Raichle, M. E. (2014). Resting state functional connectivity in early blind humans. *Front. syst. Neurosci.* 8:51.

Butt, O. H., Benson, N. C., Datta, R., and Aguirre, G. K. (2013). The fine-scale functional correlation of striate cortex in sighted and blind people. *J. Neurosci.* 33, 16209–16219. doi: 10.1523/jneurosci.0363-13.2013

Button, J. (1958). Electronics brings light to the blind. *Radio Electron* 29, 53–55.

Capelle, C., Trullemans, C., Arno, P., and Veraart, C. (1998). A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution. *IEEE Trans. Biomed. Eng.* 45, 1279–1293. doi: 10.1109/10.720206

Caraiman, S., Morar, A., Owczarek, M., Burlacu, A., Rzeszotarski, D., Botezatu, N., et al. (2017). "Computer vision for the visually impaired: the sound of vision system", in: *Proceedings of the IEEE International Conference on Computer Vision Workshops*. Piscataway, NJ; IEEE), 1480–1489.

Caspi, A., Dorn, J. D., McClure, K. H., Humayun, M. S., Greenberg, R. J., and McMahon, M. J. (2009). Feasibility study of a retinal prosthesis: spatial vision with a 16-electrode implant. *Arch. Ophthalmol.* 127, 398–401. doi: 10.1001/archophthalmol.2009.20

Castaldi, E., Cicchini, G. M., Cinelli, L., Biagi, L., Rizzo, S., and Morrone, M. C. (2016). Visual BOLD response in late blind subjects with Argus II retinal prosthesis. *PLoS Biol.* 14:e1002569. doi: 10.1371/journal.pbio.1002569

Cavaliere, C., Aiello, M., Soddu, A., Laureys, S., Reislev, N. L., Ptito, M., et al. (2020). Organization of the commissural fiber system in congenital and late-onset blindness. *NeuroImage Clin.* 25:102133. doi: 10.1016/j.nicl.2019.102133

Cecchetti, L., Kupers, R., Ptito, M., Pietrini, P., and Ricciardi, E. (2016). Are supramodality and cross-modal plasticity the yin and yang of brain development? from blindness to rehabilitation. *Front. Syst. Neurosci.* 10:89. doi: 10.3389/fnsys.2016.00089

Chebat, D. R., Chen, J. K., Schneider, F., Ptito, A., Kupers, R., and Ptito, M. (2007a). Alterations in right posterior hippocampus in early blind individuals. *Neuroreport* 18, 329–333. doi: 10.1097/wnr.0b013e32802b70f8

Chebat, D. R., Harrar, V., Kupers, R., Maidenbaum, S., Amedi, A., and Ptito, M. (2018a). "Sensory substitution and the neural correlates of navigation in blindness," in *Mobility of Visually Impaired People*. eds R. Velazquez and E. Pissaloux. Berlin: Springer), 167–200. doi: 10.1007/978-3-319-54446-5_6

Chebat, D. R., Heimler, B., Hofsetter, S., and Amedi, A. (2018b). "The implications of brain plasticity and task selectivity for visual rehabilitation of blind and visually impaired individuals," in *The Neuroimaging of Brain Diseases*. ed. C. Habas. Berlin: Springer), 295–321. doi: 10.1007/978-3-319-78926-2_13

Chebat, D. R., Maidenbaum, S., and Amedi, A. (2017). "The transfer of non-visual spatial knowledge between real and virtual mazes via sensory substitution", in *Proceedings of the 2017 International Conference on Virtual Rehabilitation (ICVR)*: Piscataway, NJ: IEEE), 1–7.

Chebat, D. R., Schneider, F. C., and Ptito, M. (2020). Spatial competence and brain plasticity in congenital blindness via sensory substitution devices. *Front. Neurosci.* 14:815. doi: 10.3389/fnins.2020.00815.

Chebat, D. R., Maidenbaum, S., and Amedi, A. (2015). Navigation using sensory substitution in real and virtual mazes. *PLoS One* 10:e0126307. doi: 10.1371/journal.pone.0126307

Chebat, D. R., Rainville, C., Kupers, R., and Ptito, M. (2007b). Tactile-'visual' acuity of the tongue in early blind individuals. *Neuroreport* 18, 1901–1904. doi: 10.1097/WNR.0b013e3282f2a63

Chebat, D. R., Schneider, F. C., Kupers, R., and Ptito, M. (2011). Navigation with a sensory substitution device in congenitally blind individuals. *Neuroreport* 22, 342–347. doi: 10.1097/WNR.0b013e3283462def

Chen, X., Wang, F., Fernandez, E., and Roelfsema, P. R. (2020). Shape perception via a high-channel-count neuroprosthesis in monkey visual cortex. *Science* 370, 1191–1196. doi: 10.1126/science.abd7435

Collignon, O., Dormal, G., Albouy, G., Vandewalle, G., Voss, P., Phillips, C., et al. (2013). Impact of blindness onset on the functional organization and the connectivity of the occipital cortex. *Brain* 136, 2769–2783. doi: 10.1093/brain/awt176

Collignon, O., Lassonde, M., Lepore, F., Bastien, D., and Veraart, C. (2007). Functional cerebral reorganization for auditory spatial processing and auditory substitution of vision in early blind subjects. *Cereb Cortex* 17, 457–465. doi: 10.1093/cercor/bhj162

Commend, S. (2001). *Les instituts Nazareth et Louis-Braille, 1861–2001: une histoire de coeur et de vision*. Les Éditions Du Septentrion (4th ed), 17–19.

d'Albe, E. F. (1920). *The Optophone: an Instrument for Reading by Ear*. Berlin: Nature Publishing Group).

De Volder, A. G., Bol, A., Blin, J., Robert, A., Arno, P., Grandin, C., et al. (1997). Brain energy metabolism in early blind subjects: neural activity in the visual cortex. *Brain Res.* 750, 235–244. doi: 10.1016/s0006-8993(96)01352-2

Delbeke, J., Pins, D., Michaux, G., Wanet-Defalque, M.-C., Parrini, S., and Veraart, C. (2001). Electrical stimulation of anterior visual pathways in retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.* 42, 291–297.

Demb, J. B., and Singer, J. H. (2015). Functional circuitry of the retina. *Annu. Rev. Vis. Sci.* 1, 263–289. doi: 10.1146/annurev-vision-082114-035334

Dresslar, F. B. J. (1893). On the pressure sense of the drum of the ear and" facial-vision. *Am. J. Psychol.* 5, 344–350. doi: 10.2307/1410997

Dufour, A., Després, O., and Candas, V. (2005). Enhanced sensitivity to echo cues in blind subjects. *Exp. Brain Res.* 165, 515–519. doi: 10.1007/s00221-005-2329-3

Edwards, T. L., Cottriall, C. L., Xue, K., Simunovic, M. P., Ramsden, J. D., Zrenner, E., et al. (2018). Assessment of the electronic retinal implant alpha AMS in restoring vision to blind patients with end-stage retinitis pigmentosa. *Ophthalmology* 125, 432–443. doi: 10.1016/j.ophtha.2017.09.019

Elli, G. V., Benetti, S., and Collignon, O. (2014). Is there a future for sensory substitution outside academic laboratories? *Multisens. Res.* 27, 271–291. doi: 10.1163/22134808-00002460

Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends Cogn. Sci.* 12, 388–396. doi: 10.1016/j.tics.2008.07.004

Farnum, A., and Pelled, G. (2020). New vision for visual prostheses. *Front. Neurosci.* 14:36. doi: 10.3389/fnins.2020.00036.

Fernandez, E. (2018). Development of visual neuroprostheses: trends and challenges. *Bioelectron. Med.* 4:12.

Fernandez, E., Alfaro, A., Toledano, R., Albisua, J., and García, A. (2015). Perceptions elicited by electrical stimulation of human visual cortex. *Invest. Ophthalmol. Vis. Sci.* 56:777.

Ferretti, G., and Glenney, B. (2020). *Molyneux's Question and the History of Philosophy*. Abingdon: Routledge.

Fortin, M., Voss, P., Lord, C., Lassonde, M., Pruessner, J., Saint-Amour, D., et al. (2008). Wayfinding in the blind: larger hippocampal volume and supranormal spatial navigation. *Brain* 131, 2995–3005. doi: 10.1093/brain/awn250

Foulke, E., Bach-Y-Rita, P., Blasch, B., Brabyn, J., Enoch, J., Faye, E., et al. (1986). "*Electronic Travel Aids: New Directions for Research*". Washington, DC: Nat. Acad. Press).

Geldard, F. A. (1957). Adventures in tactile literacy. *Am. Psychol.* 12, 115–124. doi: 10.1037/h0040416

Gulick, W. L., Gescheider, G. A., and Frisina, R. D. (1989). *Hearing: Physiological Acoustics, Neural Coding, and Psychoacoustics*. Oxford: Oxford University Press.

Guth, D., Rieser, J., and Ashmead, D. (2010). "Perceiving to move and moving to perceive: control of locomotion by students with vision loss," in *Foundations of Orientation Mobility*, 3rd ed. eds W.R. Wiener, R.L. Welsh & B.B. Blasch. New York: AFB Press), 3–44.

Heimler, B., and Amedi, A. (2020). Are critical periods reversible in the adult brain? Insights on cortical specializations based on sensory deprivation studies. *Neurosci. Biobehav. Revi.* 116, 497–507.

Heine, L., Bahri, M. A., Cavaliere, C., Soddu, A., Laureys, S., and Ptito, M. (2015). Prevalence of increases in functional connectivity in visual, somatosensory and language areas in congenital blindness. *Front. Neuroanat.* 9:86.

Heller, T. (1904). *Studien zur Blinden Psychologie*. Leipzig: W. Engelmann.

Hoffmann, R., Spagnol, S., Kristjánsson, Á., and Unnthorsson, R. (2018). Evaluation of an audio-haptic sensory substitution device for enhancing spatial awareness for the visually impaired. *Optom. Vis. Sci.* 95, 757–756. doi: 10.1097/opx.0000000000001284

Holmes, G. (1918). Disturbances of vision by cerebral lesions. *Br. J. Ophthalmol.* 2, 353–384. doi: 10.1136/bjo.2.7.353

Hoyle, B., and Waters, D. (2008). "Mobility at: the batcane (ultracane)," in *Assistive Technology for Visually Impaired and Blind People*. eds M. Hersh and M. A. Johnson. Berlin: Springer), 209–229. doi: 10.1007/978-1-84628-867-8_6

Humayun, M. S., Dorn, J. D., da Cruz, L., Dagnelie, G., Sahel, J.-A., Stanga, P. E., et al. (2012). Interim results from the international trial of second sight's visual prosthesis. *Ophthalmology* 119, 779–788. doi: 10.1016/j.ophtha.2011.09.028

Ioannides, A. A., Liu, L., Poghosyan, V., Saridis, G. A., Gjedde, A., Ptito, M., et al. (2013). MEG reveals a fast pathway from somatosensory cortex to occipital areas via posterior parietal cortex in a blind subject. *Front. Hum. Neurosci.* 7:429.

James, W. (1890). *The Principles of Psychology*. New York: Holt

Javal, É. (1905). *On Becoming Blind: Advice for the Use of Persons Losing their Sight*. London: The Macmillan company.

Jiang, J., Zhu, W., Shi, F., Liu, Y., Li, J., Qin, W., et al. (2009). Thick visual cortex in the early blind. *J. Neurosci.* 29, 2205–2211. doi: 10.1523/jneurosci.5451-08.2009

Jicol, C., Lloyd-Esenkaya, T., Proulx, M. J., Lange-Smith, S., Scheller, M., O'Neill, E., et al. (2020). Efficiency of sensory substitution devices alone and in combination with self-motion for spatial navigation in sighted and visually impaired. *Front. Psychol.* 11:1443.

Jiménez, J., Olea, J., Torres, J., Alonso, I., Harder, D., and Fischer, K.J.S. (2009). Biography of louis braille and invention of the braille alphabet. *Surv. Ophthalmol.* 54, 142–149. doi: 10.1016/j.survophthal.2008.10.006

Juurmaa, J. (1965). *An Analysis of the Components of Orientation Ability and Mental: Manipulation of Spatial Relationships*. Helsinki: Institute of Occupational Health.

Kacorri, H., Kitani, K. M., Bigham, J. P., and Asakawa, C. (2017). "People with visual impairment training personal object recognizers: feasibility and challenges", in: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. New York, NY: ACM, 5839–5849.

Kaczmarek, K., Bach-y-Rita, P., Tompkins, W. J., and Webster, J. G. (1985). A tactile vision-substitution system for the blind: computer-controlled partial image sequencing. *IEEE Trans. Biomed. Eng.* 32, 602–608. doi: 10.1109/TBME.1985.325599

Kaczmarek, K. A., Tyler, M. E., and Bach-y-Rita, P. (1994). "Electrotactile haptic display on the fingertips: preliminary results", in: *Proceedings of 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Piscataway, NJ: IEEE), 940–941.

Kay, L. (1964). An ultrasonic sensing probe as a mobility aid for the blind. *Ultrasonics* 2, 53–59. doi: 10.1016/0041-624x(64)90382-8

Kay, L. (2000). Ultrasonic eyeglasses for the blind. *J. Acoust. Soc. Am.* 108:2514. doi: 10.1121/1.4743295

Kelley, S. (2018). *Seeing AI: Artificial Intelligence for Blind and Visually Impaired Users*. Available online at: https://visionaware.org/everyday-living/helpful-products/using-apps/seeing-ai-app/

Kellogg, W. N. (1962). Sonar system of the blind. *Science* 137, 399–404.

Kendrick, D. (2005). From optacon to oblivion: the telesensory story. *Am. Foundation Blind Access World Magazine* 6.

Kohler, I. (1964). Orientation by aural clues. *J. Res. Bull. Am. Found. Blind No.* 4, 14–53.

Kupers, R., Chebat, D. R., Madsen, K. H., Paulson, O. B., and Ptito, M. (2010). Neural correlates of virtual route recognition in congenital blindness. *Proc. Natl. Acad. Sci. U S A* 107, 12716–12721. doi: 10.1073/pnas.1006199107

Kupers, R., Fumal, A., de Noordhout, A. M., Gjedde, A., Schoenen, J., and Ptito, M. (2006). Transcranial magnetic stimulation of the visual cortex induces somatotopically organized qualia in blind subjects. *Proc. Natl. Acad. Sci. U S A* 103, 13256–13260. doi: 10.1073/pnas.0602925103

Kupers, R., Pietrini, P., Ricciardi, E., and Ptito, M. (2011). The nature of consciousness in the visually deprived brain. *Front. Psychol.* 2:19.

Kupers, R., and Ptito, M. (2014). Compensatory plasticity and cross-modal reorganization following early visual deprivation. *Neurosci. Biobehav. Rev.* 41, 36–52. doi: 10.1016/j.neubiorev.2013.08.001

Lee, V. K., Nau, A. C., Laymon, C., Chan, K. C., Rosario, B. L., and Fisher, C. (2014). Successful tactile based visual sensory substitution use functions independently of visual pathway integrity. *Front. Hum. Neurosci.* 8:291. doi: 10.3389/fnhum.2014.00291

LeRoy, C. (1755). Où l'on rend compte de quelques tentatives que l'on a faites pour guérir plusieurs maladies par l'électricité. *Hist. Acad. Roy Sci. Memoires Math Phys.* 60, 87–95.

Lessard, N., Paré, M., Lepore, F., and Lassonde, M. (1998). Early-blind human subjects localize sound sources better than sighted subjects. *Nature* 395, 278–280. doi: 10.1038/26228

Levy, W. H. (1872). *Blindness and the Blind: or, a Treatise on the Science of Typhology*. London: Chapman and Hall.

Lewis, P. M., Ackland, H. M., Lowery, A. J., and Rosenfeld, J. V. (2015). Restoration of vision in blind individuals using bionic devices: a review with a focus on cortical visual prostheses. *Brain Res.* 1595, 51–73. doi: 10.1016/j.brainres.2014.11.020

Li, K. (2015). *Electronic Travel Aids for Blind Guidance—an Industry Landscape Study*. Berkeley, CA: ECS.

Linvill, J. G., and Bliss, J. C. (1966). A direct translation reading aid for the blind. *Proc. IEEE* 54, 40–51. doi: 10.1109/proc.1966.4572

Liotti, M., Ryder, K., and Woldorff, M. G. (1998). Auditory attention in the congenitally blind: where, when and what gets reorganized? *Neuroreport* 9, 1007–1012. doi: 10.1097/00001756-199804200-00010

Loomis, J. M., Klatzky, R. L., Giudice, N. A., Manduchi, R., and Kurniawan, S. (2012). Sensory substitution of vision: importance of perceptual and cognitive processing. in *Assistive Technology for Blindness and Low Vision*. eds R. Manduchi & S. Kurniawan. Boca Raton, FL: CRC Press, 162–191)

Loomis, J. M., Klatzky, R. L., Golledge, R. G., Cicinelli, J. G., Pellegrino, J. W., and Fry, P. A. (1993). Nonvisual navigation by blind and sighted: assessment of path integration ability. *J. Exp. Psychol. Gen.* 122, 73–91. doi: 10.1037/0096-3445.122.1.73

Löwenstein, K., and Borchardt, M. (1918). Symptomatologie und elektrische Reizung bei einer Schußverletzung des Hinterhauptlappens. *Deutsche Zeitschrift für Nervenheilkunde* 58, 264–292. doi: 10.1007/bf01629694

Lowery, A. J., Rosenfeld, J. V., Lewis, P. M., Browne, D., Mohan, A., Brunton, E., et al. (2015). "Restoration of vision using wireless cortical implants: the monash vision group project", in: *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2015, 1041–1044.

Lowery, A. J., Rosenfeld, J. V., Rosa, M. G. P., Brunton, E., Rajan, R., Mann, C., et al. (2017). "Monash vision group's gennaris cortical implant for vision restoration," in *Artificial Vision: a Clinical Guide*, ed. V.P. Gabel. (Cham: Springer International Publishing), 215–225. doi: 10.1007/978-3-319-41876-6_17

Luo, Y. H. L., Zhong, J., Merlini, F., Anaflous, F., Arsiero, M., Stanga, P. E., et al. (2014). The use of Argus® II retinal prosthesis to identify common objects in blind subjects with outer retinal dystrophies. *Invest. Ophthalmol. Vis. Sci.* 55, 1834–1834.

Maguire, E. A., Burgess, N., Donnett, J. G., Frackowiak, R. S., Frith, C. D., and O'Keefe, J. (1998). Knowing where and getting there: a human navigation network. *Science* 280, 921–924. doi: 10.1126/science.280.5365.921

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S., et al. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proc. Natl. Acad. Sci.* 97, 4398–4403. doi: 10.1073/pnas.070039597

Maidenbaum, S., Abboud, S., and Amedi, A. (2014a). Sensory substitution: closing the gap between basic research and widespread practical visual rehabilitation. *Neurosci. Biobehav. Rev.* 41, 3–15. doi: 10.1016/j.neubiorev.2013.11.007

Maidenbaum, S., Hanassy, S., Abboud, S., Buchs, G., Chebat, D. R., Levy-Tzedek, S., et al. (2014b). The "EyeCane", a new electronic travel aid for the blind: technology, behavior & swift learning. *Restor. Neurol. Neurosci.* 32, 813–824. doi: 10.3233/RNN-130351

Maidenbaum, S., Levy-Tzedek, S., Chebat, D. R., Namer-Furstenberg, R., and Amedi, A. (2014c). The effect of extended sensory range via the EyeCane sensory substitution device on the characteristics of visionless virtual navigation. *Multisens. Res.* 27, 379–397. doi: 10.1163/22134808-00002463

Manduchi, R., and Kurniawan, S. (2011). Mobility-related accidents experienced by people with visual impairment. *AER J. Res. Pract. Vis. Impair. Blind.* 4, 44–54.

Margo, C. E., Harman, L. E., and Smith, D. B. (2013). Blindness and the age of enlightenment: diderot's letter on the blind. *JAMA Ophthalmol.* 131, 98–102.

Matteau, I., Kupers, R., Ricciardi, E., Pietrini, P., and Ptito, M. (2010). Beyond visual, aural and haptic movement perception: hMT+ is activated by electrotactile motion stimulation of the tongue in sighted and in congenitally blind individuals. *Brain Res. Bull.* 82, 264–270. doi: 10.1016/j.brainresbull.2010.05.001

Medeiros, N. E., and Curcio, C. A. (2001). Preservation of ganglion cell layer neurons in age-related macular degeneration. *Invest. Ophthalmol. Vis. Sci.* 42, 795–803.

Meijer, P. B. (1992). An experimental system for auditory image representations. *IEEE Trans. Biomed. Eng.* 39:112. doi: 10.1109/10.121642

Morrison, C., Cutrell, E., Dhareshwar, A., Doherty, K., Thieme, A., and Taylor, A. (2017). "Imagining artificial intelligence applications with people with visual disabilities using tactile ideation", in: *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*. New York, NY: ACM, 81–90.

Muchnik, C., Efrati, M., Nemeth, E., Malin, M., and Hildesheimer, M. (1991). Central auditory skills in blind and sighted subjects. *Scand. Audiol.* 20, 19–23. doi: 10.3109/01050399109070785

Müller, F., Niso, G., Samiee, S., Ptito, M., Baillet, S., and Kupers, R. (2019). A thalamocortical pathway for fast rerouting of tactile information to occipital cortex in congenital blindness. *Nat. Commun.* 10, 1–9. doi: 10.1093/cercor/bhr363

Murphy, M. C., Nau, A. C., Fisher, C., Kim, S. G., Schuman, J. S., and Chan, K. C. (2016). Top-down influence on the visual cortex of the blind during sensory substitution. *Neuroimage* 125, 932–940. doi: 10.1016/j.neuroimage.2015.11.021

National Research Council (1986). *Electronic Travel Aids: New Directions for Research*. Washington: National Academies Press

Nau, A., Bach, M., and Fisher, C. (2013). Clinical tests of ultra-low vision used to evaluate rudimentary visual perceptions enabled by the BrainPort vision device. *Transl. Vis. Sci. Technol.* 2:1. doi: 10.1167/tvst.2.3.1

Nau, A. C., Murphy, M. C., and Chan, K. C. (2015a). Use of sensory substitution devices as a model system for investigating cross-modal neuroplasticity in humans. *Neural Regen. Res.* 10, 1717–1719. doi: 10.4103/1673-5374.169612

Nau, A. C., Pintar, C., Arnoldussen, A., and Fisher, C. (2015b). Acquisition of visual perception in blind adults using the BrainPort artificial vision device. *Am. J. Occup. Ther.* 69, 6901290010—6901290018.

Niketeghad, S., and Pouratian, N. (2019). Brain machine interfaces for vision restoration: the current state of cortical visual prosthetics. *Neurotherapeutics* 16, 134–143. doi: 10.1007/s13311-018-0660-1

Nirenberg, S., and Pandarinath, C. (2012). Retinal prosthetic strategy with the capacity to restore normal vision. *Proc. Natl. Acad. Sci.* 109, 15012–15017. doi: 10.1073/pnas.1207035109

Noppeney, U. (2007). The effects of visual deprivation on functional and structural organization of the human brain. *Neurosci. Biobehav. Rev.* 31, 1169–1180. doi: 10.1016/j.neubiorev.2007.04.012

Normann, R. A., Greger, B. A., House, P., Romero, S. F., Pelayo, F., and Fernandez, E. (2009). Toward the development of a cortically based visual neuroprosthesis. *J. Neural Eng.* 6:035001. doi: 10.1088/1741-2560/6/3/035001

Nowik, K., Langwińska-Wośko, E., Skopiński, P., Nowik, K. E., and Szaflik, J. P. (2020). Bionic eye review–an update. *J. Clin. Neurosci.* 78, 8–19. doi: 10.1016/j.jocn.2020.05.041

Pamir, Z., Canoluk, M. U., Jung, J. H., and Peli, E. (2020). Poor resolution at the back of the tongue is the bottleneck for spatial pattern recognition. *Sci. Rep.* 10, 1–13.

Paré, S., Maxime, B., Djerourou, I., Knowledge, C., Bernal, D., Piszczor, M., et al. (2019). "Sensory substitution and spatial navigation in early and late blind

individuals using a new SensoryFusion application installed on a smartphone," In *Proceedings of the Neurosymposium*. Montréal: Université de Montréal

Park, H.-J., Lee, J. D., Kim, E. Y., Park, B., Oh, M. K., Lee, S., et al. (2009). Morphological alterations in the congenital blind based on the analysis of cortical thickness and surface area. *Neuroimage* 47, 98–106. doi: 10.1016/j.neuroimage.2009.03.076

Penrod, W. M., Simmons, T. J., Bauder, D. K., and Lee, D. B. (2009). A comparison of selected secondary electronic travel aids with a primary mobility system. *Int. J. Orient Mob* 2, 27–41. doi: 10.21307/ijom-2009-003

Pezaris, J. S., and Reid, R. C. (2007). Demonstration of artificial visual percepts generated through thalamic microstimulation. *Proc. Natl. Acad. Sci.* 104, 7670–7675. doi: 10.1073/pnas.0608563104

Pissaloux, E., and Velázquez, R. (2018). "On spatial cognition and mobility strategies," in *Mobility of Visually Impaired People*. Berlin: Springer, 137–166. doi: 10.1007/978-3-319-54446-5_5

Ptito, M., and Kupers, R. (2005). Cross-modal plasticity in early blindness. *J. Integr. Neurosci.* 4, 479–488. doi: 10.1142/s0219635205000951

Ptito, M., Matteau, I., Gjedde, A., and Kupers, R. (2009). Recruitment of the middle temporal area by tactile motion in congenital blindness. *Neuroreport* 20, 543–547. doi: 10.1097/wnr.0b013e3283279909

Ptito, M., Matteau, I., Zhi Wang, A., Paulson, O. B., Siebner, H. R., and Kupers, R. (2012). Crossmodal recruitment of the ventral visual stream in congenital blindness. *Neural Plast.* 2012:304045.

Ptito, M., Moesgaard, S. M., Gjedde, A., and Kupers, R. (2005). Cross-modal plasticity revealed by electrotactile stimulation of the tongue in the congenitally blind. *Brain* 128(Pt 3), 606–614. doi: 10.1093/brain/awh380

Ptito, M., Schneider, F. C., Paulson, O. B., and Kupers, R. (2008). Alterations of the visual pathways in congenital blindness. *Exp. Brain Res.* 187, 41–49. doi: 10.1007/s00221-008-1273-4

Reich, L., Maidenbaum, S., and Amedi, A. (2012). The brain as a flexible task machine: implications for visual rehabilitation using noninvasive vs. invasive approaches. *Curr. Opin. Neurol.* 25, 86–95. doi: 10.1097/WCO.0b013e32834ed723

Reislev, N., Dyrby, T. B., Siebner, H., Lundell, H., Ptito, M., and Kupers, R. (2017). Thalamocortical connectivity and microstructural changes in congenital and late blindness. *Neural Plast.* 2017:9807512

Renier, L., Collignon, O., Poirier, C., Tranduy, D., Vanlierde, A., Bol, A., et al. (2005a). Cross-modal activation of visual cortex during depth perception using auditory substitution of vision. *Neuroimage* 26, 573–580. doi: 10.1016/j.neuroimage.2005.01.047

Renier, L., Laloyaux, C., Collignon, O., Tranduy, D., Vanlierde, A., Bruyer, R., et al. (2005b). The ponzo illusion with auditory substitution of vision in sighted and early-blind subjects. *Perception* 34, 857–867. doi: 10.1068/p5219

Röder, B., Rösler, F., Hennighausen, E., and Näcker, F. (1996). Event-related potentials during auditory and somatosensory discrimination in sighted and blind human subjects. *Cogn. Brain Res.* 4, 77–93. doi: 10.1016/0926-6410(96)00024-9

Röder, B., Teder-SaÈlejaÈrvi, W., Sterr, A., Rösler, F., Hillyard, S. A., and Neville, H. J. (1999). Improved auditory spatial tuning in blind humans. *Nature* 400(6740), 162–166. doi: 10.1038/22106

Roentgen, U. R., Gelderblom, G. J., Soede, M., and De Witte, L. P. (2008). Inventory of electronic mobility aids for persons with visual impairments: a literature review. *J. Vis. Impair. Blind.* 102, 702–724. doi: 10.1177/0145482x0810201105

Romains, J. (1924). *Eyeless sight: a Study of Extra-Retinal Vision and the Paroptic Sense*. New York and London: Putnam.

Rosenfeld, J. V., Wong, Y. T., Yan, E., Szlawski, J., Mohan, A., Clark, J., et al. (2020). Tissue response to a chronically implantable wireless, intracortical visual prosthesis (Gennaris array). *J. Neural Eng.* 4:17. doi: 10.1088/1741-2552/ab9e1c.

Russell, L. (1967). "Travel path sounder-furthe results,". in *Proceedings of the International. Conference on Sensory Devices for the Blind*, ed. R. Dufton. London: St. Dunstan's, 293–297.

Sadato, N., Okada, T., Honda, M., and Yonekura, Y. (2002). Critical period for cross-modal plasticity in blind humans: a functional MRI study. *Neuroimage* 16, 389–400. doi: 10.1006/nimg.2002.1111

Sadato, N., Okada, T., Kubota, K., and Yonekura, Y. (2004). Tactile discrimination activates the visual cortex of the recently blind naive to Braille: a functional magnetic resonance imaging study in humans. *Neurosci. Lett.* 359, 49–52. doi: 10.1016/j.neulet.2004.02.005

Sadato, N., Pascual-Leone, A., Grafman, J., Ibañez, V., Deiber, M.-P., Dold, G., et al. (1996). Activation of the primary visual cortex by Braille reading in blind subjects. *Nature* 380, 526–528. doi: 10.1038/380526a0

Sampaio, E., Maris, S., and Bach-y-Rita, P. (2001). Brain plasticity:'visual'acuity of blind persons via the tongue. *Brain Res.* 908, 204–207. doi: 10.1016/s0006-8993(01)02667-1

Santos, A., Humayun, M. S., de Juan, E. Jr., Greenburg, R. J., Marsh, M. J., Klock, I. B., et al. (1997). Preservation of the Inner retina in retinitis pigmentosa: a morphometric analysis. *JAMA Ophthalmol.* 115, 511–515. doi: 10.1001/archopht.1997.01100150513011

Sauerberger, D. (1996, May). *O&M Living History – Where did our O&M techniques come from? Newsletters of the Metropolitan Washington Orientation and Mobility Association*. Available online at: http://www.sauerburger.org/dona/omhistory.htm

Schindler, I., Rice, N. J., McIntosh, R. D., Rossetti, Y., Vighetto, A., and Milner, A. D. (2004). Automatic avoidance of obstacles is a dorsal stream function: evidence from optic ataxia. *Nat. Neurosci.* 7, 779–784. doi: 10.1038/nn1273

Shi, J., Collignon, O., Xu, L., Wang, G., Kang, Y., Leporé, F., et al. (2015). Impact of early and late visual deprivation on the structure of the corpus callosum: a study combining thickness profile with surface tensor-based morphometry. *Neuroinformatics* 13, 321–336. doi: 10.1007/s12021-014-9259-9

Shimony, J., Burton, H., Epstein, A., McLaren, D., Sun, S., and Snyder, A. (2006). Diffusion tensor imaging reveals white matter reorganization in early blind humans. *Cerebral Cortex* 16, 1653–1661. doi: 10.1093/cercor/bhj102

Shu, N., Liu, Y., Li, J., Li, Y., Yu, C., and Jiang, T. (2009). Altered anatomical network in early blindness revealed by diffusion tensor tractography. *PLoS One* 4:e7228. doi: 10.1371/journal.pone.0007228

Smith, D., and Penrod, W. (2010). "Adaptive technology for orientation and mobility," in *Foundations of Orientationand Mobility*, eds W. Wiener, R.L. Welsh & B.B. Blasch. (New York: AFB Press), 241–276.

Squire, L. R. (2009). Memory and brain systems: 1969–2009. *J. Neurosci.* 29, 12711–12716. doi: 10.1523/jneurosci.3575-09.2009

Stein, D. (1998). *The Optacon: Past, Present, and Future*. Available online at: http://www.nfb.org/bm/bm98/bm980506.htm (accessed November 18, 2020).

Stingl, K., Bartz-Schmidt, K. U., Besch, D., Braun, A., Bruckmann, A., Gekeler, F., et al. (2013). Artificial vision with wirelessly powered subretinal electronic implant alpha-IMS. *Proc. R. Soc. B Biol. Sci.* 280:20130077. doi: 10.1098/rspb.2013.0077

Strelow, E. R., and Brabyn, J. A. (1982). Locomotion of the blind controlled by natural sound cues. *Perception* 11, 635–640. doi: 10.1068/p110635

Striem-Amit, E., and Amedi, A. (2014). Visual cortex extrastriate body-selective area activation in congenitally blind people "seeing" by using sounds. *Curr. Biol.* 24, 687–692. doi: 10.1016/j.cub.2014.02.010

Striem-Amit, E., Cohen, L., Dehaene, S., and Amedi, A. (2012a). Reading with sounds: sensory substitution selectively activates the visual word form area in the blind. *Neuron* 76, 640–652. doi: 10.1016/j.neuron.2012.08.026

Striem-Amit, E., Guendelman, M., and Amedi, A. (2012b). 'Visual' acuity of the congenitally blind using visual-to-auditory sensory substitution. *PLoS One* 7:e33136. doi: 10.1371/journal.pone.0033136

Stronks, H. C., and Dagnelie, G. (2014). The functional performance of the Argus II retinal prosthesis. *Expert Rev. Med. Devices* 11, 23–30. doi: 10.1586/17434440.2014.862494

Stronks, H. C., Mitchell, E. B., Nau, A. C., and Barnes, N. (2016). Visual task performance in the blind with the BrainPort V100 Vision Aid. *Expert Rev. Med. Devices* 13, 919–931. doi: 10.1080/17434440.2016.1237287

Supa, M., Cotzin, M., and Dallenbach, K. M. (1944). " Facial vision": the perception of obstacles by the blind. *Am. J. Psychol.* 57, 133–183. doi: 10.2307/1416946

Suterko, S. (1967). "Long cane training: Its advantages and problems", in: *Proceedings of the Conference for Mobility Trainers and Technologists*. Cambridge, MA: Massachusetts Institute of Technology, 13–18.

Thaler, L., and Goodale, M. A. (2016). Echolocation in humans: an overview. *J. Wiley Interdisciplinary Rev. Cogn. Sci.* 7, 382–393. doi: 10.1002/wcs.1408

Tomaiuolo, F., Campana, S., Collins, D. L., Fonov, V. S., Ricciardi, E., Sartori, G., et al. (2014). Morphometric changes of the corpus callosum in congenital blindness. *PLoS One* 9:e107871. doi: 10.1371/journal.pone.0107871

Troyk, P. R. (2017). "The intracortical visual prosthesis project," in *Artificial Vision: a Clinical Guide*, ed. V. P. Gabel. (Cham: Springer International Publishing), 203–214. doi: 10.1007/978-3-319-41876-6_16

Truschel, L. (1906). Dersechste sinn der blinden [the sixth sense of the blind]. *Blinden. Exp. Pad.* 3, 109–142.

Villey, P. (1930). *The world of the Blind; a Psychological Study*. New York: Macmillan.

Vincent, M., Tang, H., Zhu, Z., and Ro, T. (2014). Discrimination of shapes and line orientations on the tongue. *J. Vis.* 14, 1094–1094. doi: 10.1167/14.10.1094

Voss, P., Lassonde, M., Gougoux, F., Fortin, M., Guillemot, J.-P., and Lepore, F. (2004). Early-and late-onset blind individuals show supra-normal auditory abilities in far-space. *Curr. Biol.* 14, 1734–1738. doi: 10.1016/j.cub.2004.09.051

Wan, C. Y., Wood, A. G., Reutens, D. C., and Wilson, S. J. (2010). Early but not late-blindness leads to enhanced auditory perception. *Neuropsychologia* 48, 344–348. doi: 10.1016/j.neuropsychologia.2009.08.016

Watkins, K. E., Shakespeare, T. J., O'Donoghue, M. C., Alexander, I., Ragge, N., Cowey, A., et al. (2013). Early auditory processing in area V5/MT+ of the congenitally blind brain. *J. Neurosci.* 33, 18242–18246. doi: 10.1523/jneurosci.2546-13.2013

Wen, Z., Zhou, F.-Q., Huang, X., Dan, H. D., Xie, B.-J., and Shen, Y. (2018). Altered functional connectivity of primary visual cortex in late blindness. *Neuropsychiatr. Dis. Treat.* 14, 3317–3327. doi: 10.2147/ndt.s183751

WeWALK (2019). *WeWALK Smart Cane [Online]*. Available online at: https://wewalk.io/en/ [accessed November 20 2020].

Weygand, Z. (2009). The blind in french society: images and institutions from the middle ages to the 19th century. *Revue d'eìthique et de theìologie morale* 256**:**65. doi: 10.3917/retm.256.0065.

Wheatley, E. (2010). *Stumbling blocks before the blind: medieval constructions of a disability*. Michigan: University of Michigan Press.

Wiley, J., Olzak, L., and Thomas, J. (1986). *Handbook of Perception and Human Performance. Volume 1: Sensory Processes and Perception. Chapter 7: Seeing Spatial Patterns*. Los Angeles: University of California.

Williams, M. D., Ray, C. T., Griffith, J., and De l'Aune, W. (2011). The use of a tactile-vision sensory substitution system as an augmentative tool for individuals with visual impairments. *J. Vis. Impair. Blind.* 105, 45–50. doi: 10.1177/0145482x1110500105

Wittenberg, G. F., Werhahn, K. J., Wassermann, E. M., Herscovitch, P., and Cohen, L. G. (2004). Functional connectivity between somatosensory and visual cortex in early blind humans. *Eur. J. Neurosci.* 20, 1923–1927. doi: 10.1111/j.1460-9568.2004.03630.x

Worchel, P., and Ammons, C. (1953). The course of learning in the perception of obstacles. *Am. J. Psychol.* 41, 170–176

Worchel, P., and Dallenbach, K. M. (1947). " Facial vision:" perception of obstacles by the deaf-blind. Am. J. Psychol. 60, 502–553. doi: 10.2307/1417725

Worchel, P., Mauney, J., & Andrew, J. G. (1950). The perception of obstacles by the blind. *J. Exp. Psychol.* 40, 746–751. doi: 10.1037/h0060950

Yanai, D., Weiland, J. D., Mahadevappa, M., Greenberg, R. J., Fine, I., and Humayun, M. S. (2007). Visual performance using a retinal prosthesis in three subjects with retinitis pigmentosa. *Am. J. Ophthalmol.* 143, 820–827.e822. doi: 10.1016/j.ajo.2007.01.027.

Zhao, Y., Wu, S., Reynolds, L., and Azenkot, S. (2018). "A face recognition application for people with visual impairments: understanding use beyond the lab", in: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. New York, NY: ACM, 1–14.

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

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 01 April 2021
doi: 10.3389/fnhum.2021.639871

Clinical Recognition of Sensory Ataxia and Cerebellar Ataxia

Qing Zhang[1,2], Xihui Zhou[1], Yajun Li[2], Xiaodong Yang[3] and Qammer H. Abbasi[4]*

[1] First Affiliated Hospital of Xi'an Jiaotong University, Xi'an Jiaotong University Health Science Center, Xi'an Jiaotong University, Xi'an, China, [2] Northwest Women's and Children's Hospital, Xi'an Jiaotong University Health Science Center, Xi'an, China, [3] School of Electronic Engineering, Xidian University, Xi'an, China, [4] James Watt School of Engineering, University of Glasgow, Glasgow, United Kingdom

Ataxia is a kind of external characteristics when the human body has poor coordination and balance disorder, it often indicates diseases in certain parts of the body. Many internal factors may causing ataxia; currently, observed external characteristics, combined with Doctor's personal clinical experience play main roles in diagnosing ataxia. In this situation, different kinds of diseases may be confused, leading to the delay in treatment and recovery. Modern high precision medical instruments would provide better accuracy but the economic cost is a non-negligible factor. In this paper, novel non-contact sensing technique is used to detect and distinguish sensory ataxia and cerebellar ataxia. Firstly, Romberg's test and gait analysis data are collected by the microwave sensing platform; then, after some preprocessing, some machine learning approaches have been applied to train the models. For Romberg's test, time domain features are considered, the accuracy of all the three algorithms are higher than 96%; for gait detection, Principal Component Analysis (PCA) is used for dimensionality reduction, and the accuracies of Back Propagation (BP) neural Network, Support Vector Machine (SVM), and Random Forest (RF) are 97.8, 98.9, and 91.1%, respectively.

Keywords: cerebellar ataxia, clinical recognition, microwave, sensory ataxia, wireless sensing technology

OPEN ACCESS

Edited by:
Victor Hugo C. de Albuquerque, University of Fortaleza, Brazil

Reviewed by:
Daniyal Haider, De Montfort University, United Kingdom
Hassan Chattha, Okanagan College, Canada

****Correspondence:***
Xiaodong Yang
xdyang@xidian.edu.cn

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

Received: *10 December 2020*
Accepted: *27 January 2021*
Published: *01 April 2021*

Citation:
Zhang Q, Zhou X, Li Y, Yang X and Abbasi QH (2021) Clinical Recognition of Sensory Ataxia and Cerebellar Ataxia. Front. Hum. Neurosci. 15:639871. doi: 10.3389/fnhum.2021.639871

INTRODUCTION

"Ataxia" was initially used to describe various uncoordinated characteristics of different diseases, such as gait, movement, heartbeat, etc. Now it is more specifically used to express the symptoms of motor mismatching synchronization and balance disorder after the brain, cerebellum, deep sensation (proprioception), vestibular and other systems are damaged (Bastian, 1997). Different pathological locations often show different characteristics. Sensory ataxia is caused by the impairment of somatosensory nerve, which leads to the interruption of sensory feedback signals and therefore, the body incoordination is caused. For Cerebellar Ataxia patients, the Romberg's sign was positive, the typical symptoms include walking slowly, rolling, etc. Symptoms were mild when eyes were open and aggravated when eyes were closed (Fadic et al., 1997; Donnelly, 2011). Cerebellar ataxia patients are more common, it is a loss of body muscle coordination caused by cerebellar disease. Trunk ataxia often indicates cerebellar vermis lesions, and limb ataxia often indicates cerebellar hemisphere lesions. The corresponding patients often have eye tremor, low muscle tension, unclear speech, and other symptoms (Diener and Dichgans, 1992; Bastian et al., 1996).

In clinical testing, SA syndrome is very easy to be misdiagnosed as CA syndrome, which leads to the inability of patients with ataxia to get correct diagnosis and treatment in time. At present, several international medical organizations have formed to study ataxia (Klockgether and Paulson, 2011), and several ataxia assessment scales were developed, such as "International Cooperative Ataxia Rating Scale for pharmacological assessment of the cerebella syndrome (ICARS)" (Trouillas et al., 1997), "Scale for the assessment and rating of ataxia (SARA)" (Schmitz-Hübsch, 2006). Some scholars have also done relevant research on the clinical detection and differentiation of SA and CA symptoms, and given the clinical diagnosis method (Chhetri et al., 2014). Both the assessment scale and related research work have referred to two basic indicators: Romberg's sign and gait; which could be used for Clinical detection and differentiation of SA and CA.

Romberg's Sign

The maintenance of human balance mainly depends on the coordination of vestibular system, visual system and proprioceptive system (Maurer et al., 2001). In an upright position, a normal person can stand steadily when the eyes open and close; but when two or more systems are damaged, the human body will not be able to maintain balance. For example, when a patient is suffering from Sensory Ataxia, the visual system can provide compensation information when the eyes are open, so the patient can remain upright and stable; Visual compensation would disappear when the eyes are closed, patients will not be able to maintain upright stability. This is the theoretical basis of Romberg's sign has become an important part of modern neurological clinical examination (Lanska, 2002).

In Romberg's test, the patient's feet are closed and arms are placed on both sides of the body. Standing is divided into two stages: opening eyes and closing eyes. Firstly, the patients are allowed to open their eyes and stand for a certain time, then the patients close the eyes and stand for a while, and the patients are observed: whether their body have obvious shaking in two stages. As long as there is a stage in which the patient shows standing instability, the Romberg's sign is positive (Pearce, 2005). Before carrying out Romberg's test, lower limb diseases or other factors should be excluded. In order to prevent the patient from falling down, protective pads should be laid around the patient's standing and medical staff should also take care of the patients. During the experiment, normal people can keep their body stable whether they open or close their eyes. Considering age, gender and other factors, the normal performance of the minimum standard should be that body balance can be maintained for 6 s during eye closure (Hain and Cherchi, 2017). For sensory ataxia and cerebellar ataxia, their Romberg's signs are both positive, but there are some differences. The patient can keep standing steady during the eye-opening phase, and standing unsteadily, wobbling, or even falling in the closed eye phase (Franchignoni et al., 1984), as shown in **Figure 1**. The cerebellar ataxia patients were unstable in the stage of closing eyes and opening eyes, and tend to tilt toward the diseased side of cerebellum (Cazzato et al., 2016), as shown in **Figure 2**. Romberg's test is a simple and sensitive clinical trial, the different performances of normal people, sensory ataxia patients and cerebellar ataxia patients in the Romberg's test are given in **Table 1**.

Gait Detection

Abnormal gait can be caused by motor or sensory disturbance, and its characteristics are related to the location of lesion. It can be seen in many diseases in nervous and other systems; some typical abnormal gaits have implications for certain diseases (Thomann and Dul, 1996).

Sensory Ataxia Gait

When a normal person walks, the sensory nerve would be stimulated when the sole of the foot touches the ground, then the relevant information is transmitted to indicate the position of the feet. Since the patients with sensory ataxia lose the input of the stimulus, in order to know the time and place the feet land, the patient would put his feet on the ground heavily. The key to this gait is that when patients can't see their feet (e.g., in the dark), stepping will increase obviously. This gait is sometimes referred to as stepping gait, because patients may lift their legs to a very high position (Missaoui et al., 2013). The sensory ataxia gait diagram is shown in **Figure 3**.

Cerebellar Ataxia Gait

This gait is common in cerebellar diseases and is often described as a clumsy, tottering, and wide-base gait. Similar to the gait after acute alcoholism, patients will not be able to walk straight. Patients with greater trunk instability during walking are more likely to have lesions in the midline vermis of the cerebellum (Mochizuki and Ugawa, 2010). The cerebellar ataxia gait diagram is shown in **Figure 4**.

At present, there are many related works on quantifying the degree of swing in Romberg's test.

One of the common ways is wearing facilities such as pressure sensor, gravity acceleration sensor, etc. (Diener et al., 1984; Lanska, 2002; Mcgough et al., 2018); other ways include collecting videos via cameras (Havasi et al., 2007), etc. Currently, many related works have been done for Romberg's test and gait detection purpose. Pressure sensors, gravity acceleration sensors, videos, and some other approaches have been applied in this domain (Diener et al., 1984; Lanska, 2002; Zongyi and Sarkar, 2006; Havasi et al., 2007; Afendi et al., 2013; Umair Bin Altaf et al., 2015; Wang et al., 2017; Mcgough et al., 2018). The methods in previous work have their respective advantages; however, some issues like self-consciousness enhancing, abnormal mood changes cannot be ignored. Non-contact wireless sensing technology could avoid these problems and by using omnidirectional antennas, Romberg's test and gait detection can be achieved.

The steps can be summarized as follows: firstly, the microwave sensing system working at 4.8GHz was used to collect original perception data; then, the data were preprocessed; finally, the features are extracted and three machine learning algorithms [Back Propagation (BP) Neural Network (Rumelhart et al., 1995), Support Vector Machine (SVM) (Cortes and Vapnik, 1995) and Random Forest (RF) (Shi and Horvath, 2006)] were applied to train the models. The experimental results show that

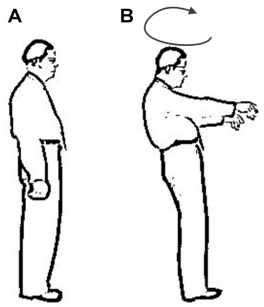

FIGURE 1 | Comparison of different performance of patients with sensory ataxia in Romberg's test. **(A)** Eye opening stage and **(B)** eye closure stage.

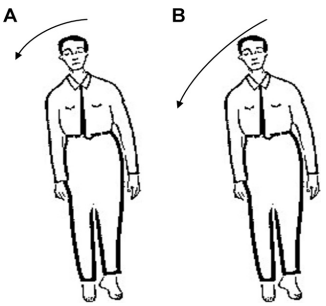

FIGURE 2 | Comparison of different performance of patients with cerebellar ataxia in Romberg's test. **(A)** Eye opening stage and **(B)** eye closure stage.

TABLE 1 | Performance of different groups in Romberg's test.

Groups	Open eyes stage	Close eyes stage
Normal	Stable	Stable
Sensory ataxia	Stable	Instable
Cerebellar ataxia	Instable	Instable

FIGURE 3 | Sensory ataxia Gait.

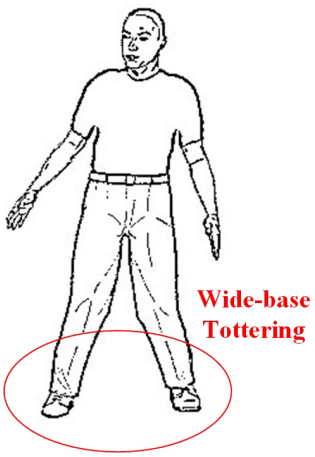

FIGURE 4 | Cerebellar ataxia Gait.

the accuracies of three algorithms are higher than 96% for Roberg's test and gait detection, demonstrating the feasibility and effectiveness of the method.

The contribution of this paper can be summarized as follows: (1) detection and distinguishing of sensory ataxia and cerebellar ataxia can be achieved by using wireless sensing technology, and the patients' privacy can be protected; (2) Romberg's test and gait detection are validated, thus the accuracy of clinical diagnosis can be improved; (3) various machine learning algorithms are used to increase the stability and credibility of the results.

The rest of the paper is organized as follows: the principle of wireless sensing is introduced in part II; in part III, the experimental devices and scheme are described in detail; in part IV, the data for Romberg's test and gait detection are analyzed; and the experimental results are discussed in part V; and in part VI, the full paper is summarized.

PRINCIPLE OF C-BAND WIRELESS SENSORY

In typical indoor environment, the wireless signal emitted by the transmitter would be affected by the objects or the human body; and the refraction, reflection and diffraction may cause multipath effect. These homologous wireless signals in different propagation paths show different physical characteristics at the receiving end, such as the amplitude and phase of the receiving signals, which contain rich information from the external environment.

When the receiver detects that the signal changes, it indicates that the external environment has been changed. By de-noising the acquired data and further processing with classification algorithm, we can reduce the environmental factors that lead to the change of the received signal, so as to obtain

the desirable information in the environment. In this work, since the antennas were used for detection and monitoring applications in a regular shape room, basic omnidirectional monopole antennas were considered; for irregular shape space and room, specially designed antennas would be necessary to enhance the performance and accuracy of sensing. The main differences between received signal strength and channel state information are explained in Zhu and Zhang (2010) and Zheng et al. (2013). CSI considers the number of antennas and subcarriers, and can measure more fine-grained information, the facility which confirms to the IEEE 802.11n standard was used to collect the CSI data. The IEEE 802.11n standard uses orthogonal frequency division multiplexing (OFDM) to transmit a single data stream with 20 MHz bandwidth through 56 orthogonal subcarriers, the signals transmitted on each subcarrier have different signal strength and phase (Lorincz and Begusic, 2006). The facility used in this paper provides 30 available subcarriers to users. Next, we will further explain the principle of C-Band wireless sensing measurement from the formula.

It is known that the channel impulse response (CIR) is generally used to describe the multipath effect in wireless channels. Under linear time-invariant conditions, the CIR can be expressed as follows,

$$h(\tau) = \sum_{i=1}^{N} a_i e^{-j\theta_i} \delta(\tau - \tau_i) \quad (i = 1, 2, \ldots, N) \tag{1}$$

In the formulas above, a_i, θ_i and τ_i represent the attenuation factor, phase shift, and time delay of the i-th path, respectively, N is the total number of propagation paths, and $\delta(\tau)$ is Dirichlet pulse function.

Since the multipath propagation of signals can cause delay and attenuation, we can also describe the channel by channel frequency response (CFR), as shown in (2),

$$Y = HX + N(2) \tag{2}$$

Where Y is the vector representation of receiving signal, X is the vector representation of transmitting signal, N is the noise matrix, H is the channel attenuation matrix and describes the attenuation factor of signal on each transmission path, the dimension of H can be expressed as:

$$Dim_H = R_N \times T_N \times Sub_N \tag{3}$$

Where R_N and T_N are the number of receiving antennas and transmitting antennas, respectively.Sub_N is the number of subcarriers.

CSI is essentially a representation of the frequency response of each subcarrier channel, as shown in (4),

$$h(f_i, t) = |h(f_i, t)| \times \arg(h(f_i, t)) \; (i = 1, 2, \ldots, 30) \tag{4}$$

In (4),$|h(f_i, t)|$, arg(h (f_i, t)), and f_i denote the amplitude, phase, and central frequency of i-th subcarrier, respectively.

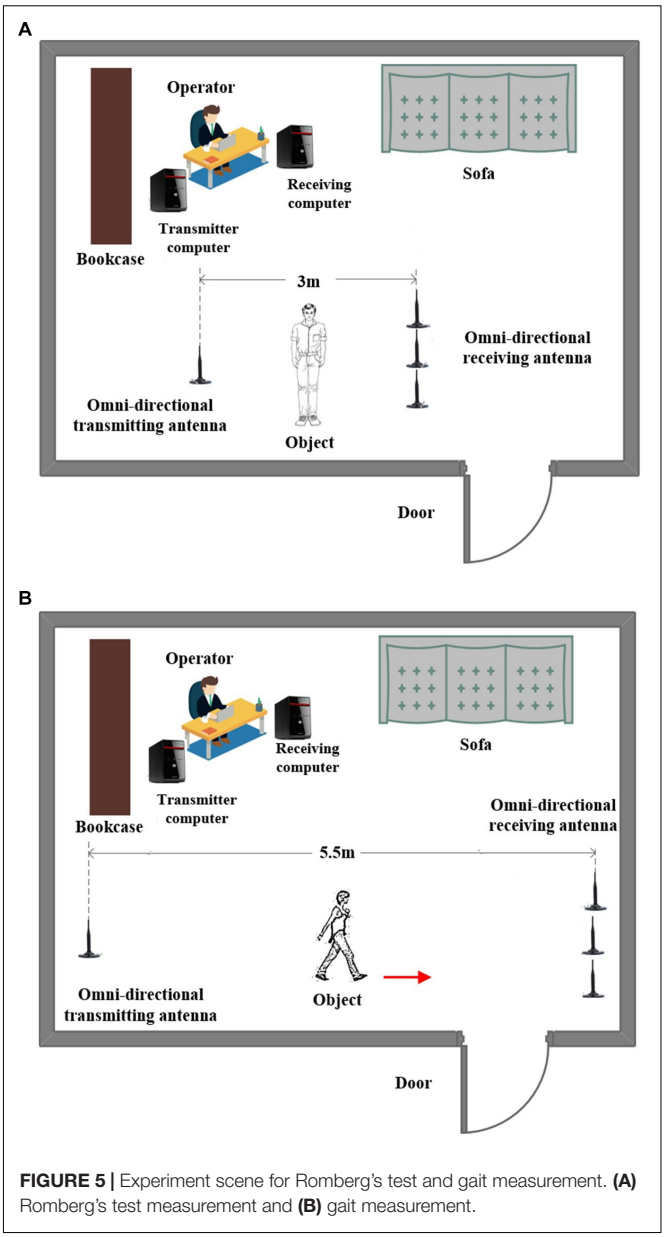

FIGURE 5 | Experiment scene for Romberg's test and gait measurement. **(A)** Romberg's test measurement and **(B)** gait measurement.

Since the patient takes some time to perform Romberg's test and gait detection, we need continuous monitoring, and the received CSI data can be expressed as:

$$D = [P_1, P_2, \ldots, P_n] \tag{5}$$

Where D represents the data stream received by the receiving antenna, P_i (i = 1, 2..., n) represents packet. Each packet contains 30 subcarriers, and n is the total number of received packets. D constitutes the analysis data source for detecting and distinguishing sensory ataxia and cerebellar ataxia. Since the phase of subcarriers in each packet is random, this paper will mainly use the amplitude information of subcarriers (Yang et al., 2017).

THE EXPERIMENT DESIGN

The experiment was carried out in an approximate ward room, its size is 7 m× 5 m.

The experimental equipment includes two industrial control computers equipped with facilities conforming to the IEEE 802.11n standard. The transmitter is equipped with an omnidirectional antenna, and the receiver is equipped with three omnidirectional antennas. Since each antenna receives packets containing 30 subcarriers, we will get 3× 30 subcarriers for each packet at the receiving end, which greatly increases the data size. The experimental scene is shown in **Figure 5**.

In the experiment, 10 subjects are considered; they are divided into two groups with five people in each group. We set the contract awarding frequency to 200 Hz. For Romberg's test, we collected a total of 12 s of data, including 6 s for the open eyes stage and 6 s for the close eyes stage. For the gait detection experiment, considering the site constraints and the walking speed between different objects, we collected a total of 5 s of data, the amount of data is enough to distinguish the abnormal gait.

For each subject, Romberg's test and gait detection were repeated 24 times. We collected three sets of data each day and collected the complete data in about 1 week. There are 120 sets of experimental data for each of the test items for sensory ataxia and cerebellar ataxia. At the same time, we also collected 120 sets of Romberg's test and gait detection data under normal conditions as a reference.

THE DATA PROCESSING

Due to the noise in the environment, to ensure the credibility and accuracy of the results, the data is processed considering the following steps (**Figure 6**):

Data Preprocessing

Remove Outliers

In order to explain the method of removing outliers, we randomly select a group of original experimental data from normal person in Romberg's test, and randomly select a subcarrier (No. 27). The signal curve of the subcarrier is shown in **Figure 7A**. When a normal person performs Romberg's test, the body shake is within a certain range, and the signal curve of subcarrier should be relatively stable, but in **Figure 7B**, there are many burrs in the signal curve and the volatility is large. We could also use the Hampel function based on the Pauta criterion to complete the removal of the outliers in the original signal (Li et al., 2016).

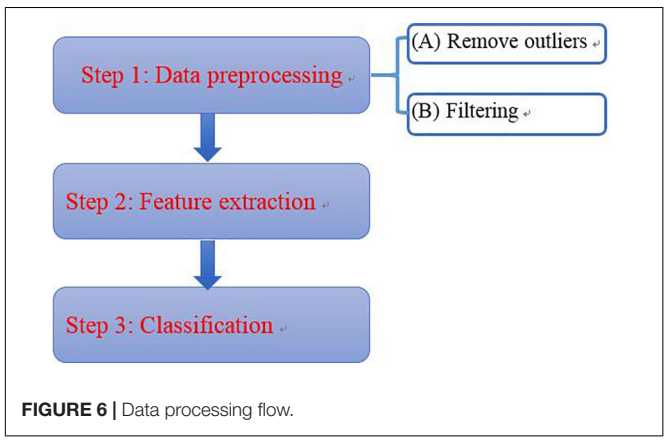

FIGURE 6 | Data processing flow.

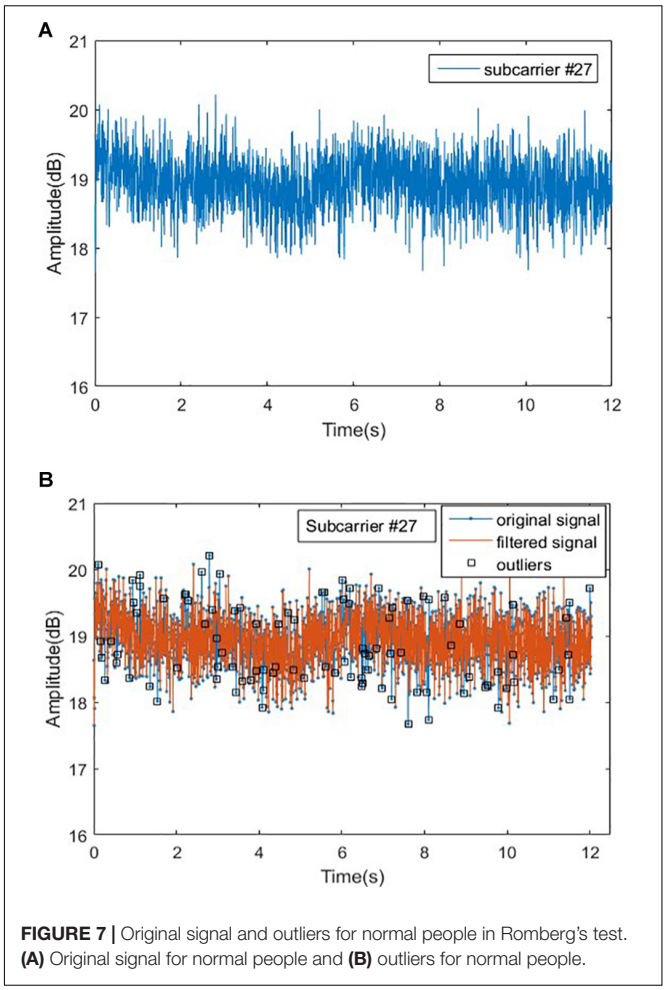

FIGURE 7 | Original signal and outliers for normal people in Romberg's test. **(A)** Original signal for normal people and **(B)** outliers for normal people.

Signal Denoising

After removing the outliers from the original signal, the noise contained in the original signal will be filtered out. Conventional filters mainly include linear filters and nonlinear filters such as mean filter and Wiener filter. The shortcoming of the traditional denoising method is that the entropy after signal transformation would increase, the non-stationary characteristics of the signal cannot be characterized, and the correlation of the signal cannot be obtained. To overcome these issues, the wavelet transform is used.

Wavelet transform has the characteristics of low entropy, multi-resolution, and flexible selection of wavelet basis functions. In this paper, the wavelet soft threshold method is used for signal denoising, which is simple to implement, and very suitable for processing low SNR (Poornachandra, 2008). We denoise the signal according to the following steps: (1)

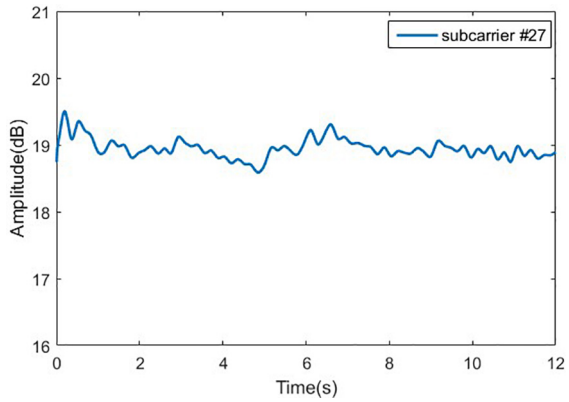

FIGURE 8 | Wavelet transform filtering effect.

Wavelet decomposition; (2) Threshold quantization of high-frequency coefficients of wavelet decomposition; and (3) Wavelet reconstruction. The wavelet function selected in this work is sym8, and the signal is decomposed into 5 layers. At the same time, in step 2, the threshold is dynamically adjusted according to the noise level of different decomposition layers. The experimental results show that the wavelet transform has smooth denoising effect, which is shown in **Figure 8**.

Feature Extraction

Select Subcarrier

Before feature extraction, it is necessary to pick out the appropriate subcarriers. We know that when the variance of a set of data is larger, more information will be contained. According to the principle of maximum variance, for Romberg's test, we select the 10th subcarrier of the third antenna; and for the gait detection experiment, we select the 26th subcarrier of the second antenna. The experimental data of selected subcarriers are shown in **Figures 9**, **10**, respectively.

Extracting Feature of the Romberg's Test Data

As seen in **Figure 9C**, in Romberg's test, the normal person can maintain balance even if he blinks or closes his eyes; slight fluctuations might be caused by the breathing of the object and the noise in the environment. Patients with sensory ataxia can maintain body balance during the blinking phase, and the body violently shakes during the closed eyes stage, resulting in a waveform that is basically stable in the blinking phase and unstable in the closed eyes phase, as shown in **Figure 9A**. For patients with cerebellar ataxia, whether they are blinking or closing their eyes, the body is shaking sharply, and the waveform fluctuates sharply, as shown in **Figure 9B**.

Since different groups in the Romberg's test have different time domain waveforms, in order to improve the efficiency of the classification model training, only the time domain characteristics are extracted and are shown in **Table 2**.

The physical significance of each time domain feature is as follows: the mean value describes the stable component of the signal, the mean square value reflects the energy of the signal,

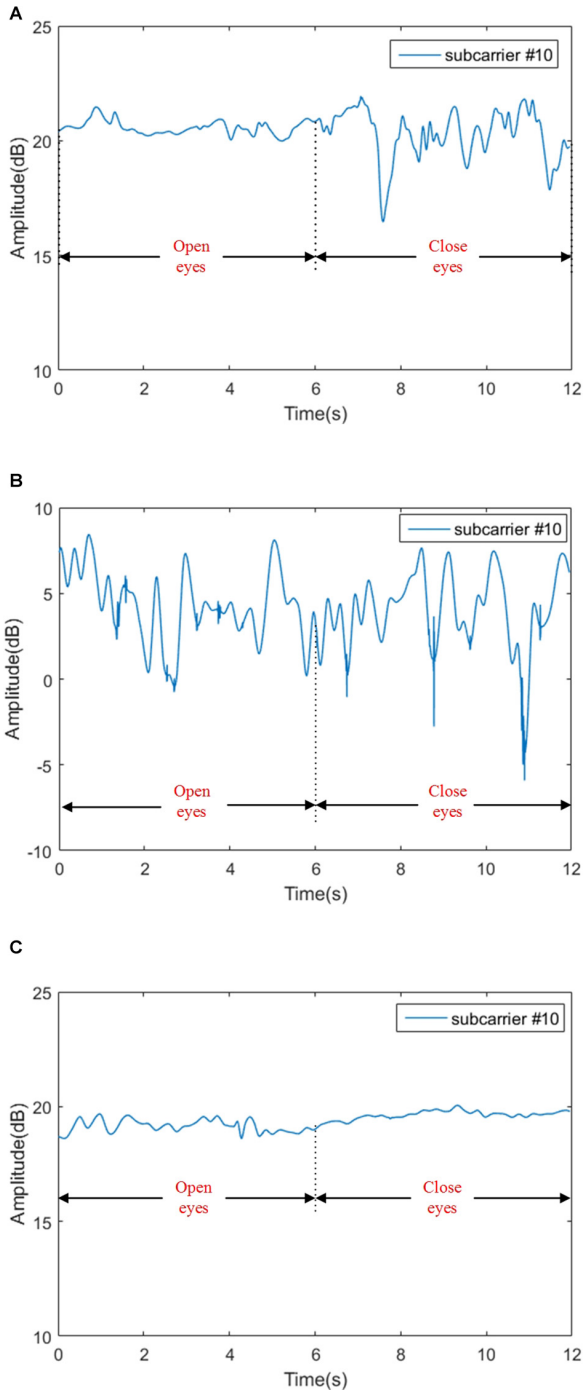

FIGURE 9 | Signal amplitudes for sensory ataxia subject, cerebellar ataxia subject and normal person in Romberg's test. **(A)** Signal amplitudes for sensory ataxia subject, **(B)** signal amplitudes for cerebellar ataxia subject, and **(C)** signal amplitudes for normal person.

the standard deviation can represent the degree of dispersion between the signal sampling points, the kurtosis reflects the impact characteristics in the signal, and the skewness reflects the asymmetry of the signal. The peak-to-peak value reflects the

TABLE 2 | The extracted time domain features of the Romberg's test data.

Features	Calculation formula
Mean value	$Y_{MV} = \frac{1}{N}\sum_{i=1}^{N} x_i$
Standard deviation	$Y_{SD} = \sqrt[2]{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - Y_{MV})^2}$
Root mean square	$Y_{RMS} = \sqrt[2]{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$
Peak to peak value	$Y_{PPV} = \max(x_i) - \min(x_i)(i = 1, 2, \ldots, N)$
Kurtosis	$Y_K = \frac{\frac{1}{N}\sum_{i=1}^{N}(\lvert x_i\rvert - Y_{MV})^4}{{Y_{RMS}}^4}$
Skewness	$Y_S = \frac{\frac{1}{N}\sum_{i=1}^{N}(\lvert x_i\rvert - Y_{MV})^3}{{Y_{RMS}}^3}$
Peak factor	$Y_P = \frac{\max(x_i)}{Y_{RMS}}\ (i = 1, 2, \ldots, N)$
Waveform factor	$Y_W = \frac{N * Y_{RMS}}{\sum_{i=1}^{N}\lvert x_i\rvert}\ (i = 1, 2, \ldots, N)$

signal amplitude range. The peak factor can be used to detect whether there is an impact in the signal. The physical meaning of the waveform factor in the electronic field can be understood as the ratio of the DC signal of the same power to the original AC signal, and its value is greater than or equal to 1.

Extracting Feature of the Gait Detection Data

As it can be seen in **Figure 10**, the time domain waveforms of the three gaits have little discrimination. To ensure the accuracy of the results, Principal Component Analysis (PCA) (Wold et al., 1987) is adopted to reduce the dimensionality of the original data, and the cumulative contribution rate of each principal component is shown in **Figure 11**.

In order to avoid information loss in the original data and to eliminate redundant information, the first 64 principal components are extracted as features.

Classification

After these steps, we have obtained the dataset of Romberg's test and gait detection. Each dataset contains 360 samples, including normal, sensory ataxia and cerebellar ataxia. To increase the credibility and accuracy of the results, we adopted a four-fold cross-validation (Demsar, 2006) method to divide the training set and test set, and adopted three classification algorithms including BP Neural Network, SVM, and RF.

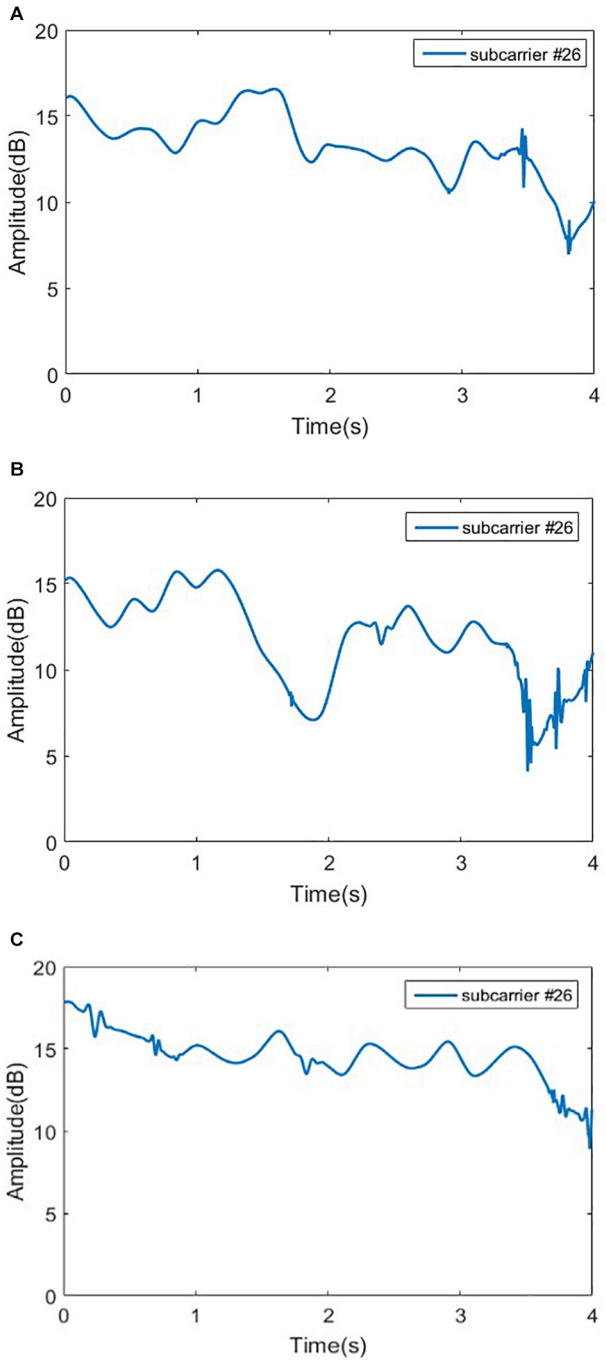

FIGURE 10 | Signal amplitudes for sensory ataxia subject, cerebellar ataxia subject and normal person in gait test. **(A)** Signal amplitudes for sensory ataxia subject, **(B)** signal amplitudes for cerebellar ataxia subject, and **(C)** signal amplitudes for normal person.

EXPERIMENTAL RESULTS AND DISCUSSION

Experimental Results

The confusion matrix for the results are shown in **Tables 3**, **4**, and the accuracies of each algorithm is shown in **Figures 12**, **13**.

Discussion

It can be seen from **Figures 12**, **13** that for Romberg's test, only the time domain features are extracted, and all the three algorithms can achieve an accuracy of more than 96%; for gait detection, PCA is used for dimensionality reduction; the accuracies of

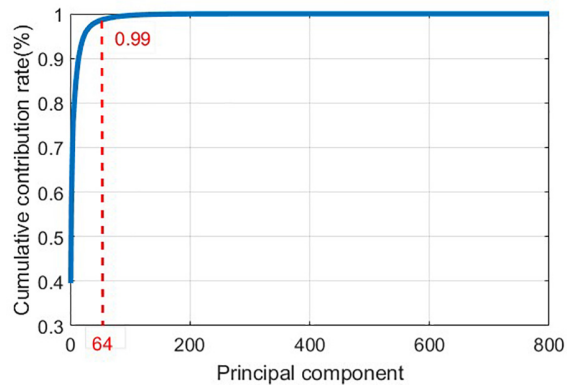

FIGURE 11 | Principal component cumulative contribution rate curve.

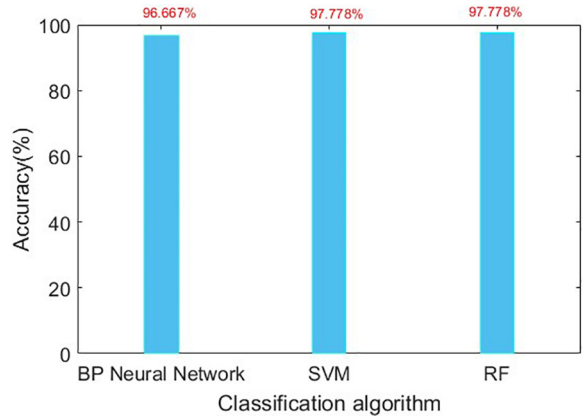

FIGURE 12 | Accuracy of three algorithms in Romberg's test.

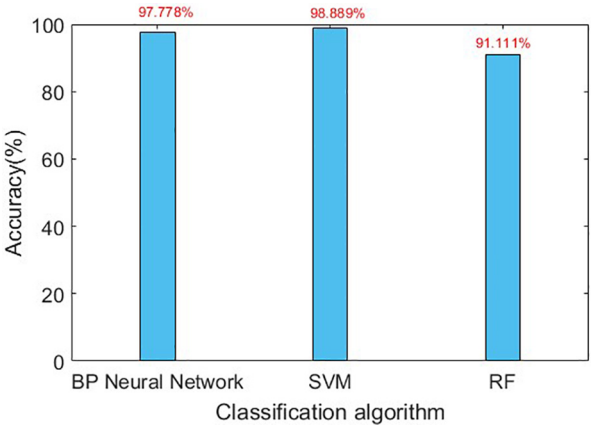

FIGURE 13 | Accuracy of three algorithms in gait detection.

TABLE 3 | Confusion matrix for Romberg's test.

Classification algorithm	Actual type (Each test set contains 3*30 samples)	Predict type (Number of sample)		
		Normal	Sensory ataxia	Cerebellar ataxia
BP Neural Network	Normal	29	1	0
	Sensory ataxia	2	28	0
	Cerebellar ataxia	0	0	30
SVM	Normal	29	1	0
	Sensory ataxia	1	29	0
	Cerebellar ataxia	0	0	30
RF	Normal	29	1	0
	Sensory ataxia	1	29	0
	Cerebellar ataxia	0	0	30

TABLE 4 | Confusion matrix for gait detection.

Classification algorithm	Actual type (Each test set contains 3*30 samples)	Predict type (Number of sample)		
		Normal	Sensory ataxia	Cerebellar ataxia
BP Neural Network	Normal	30	0	0
	Sensory ataxia	0	28	2
	Cerebellar ataxia	0	0	30
SVM	Normal	29	0	1
	Sensory ataxia	0	30	0
	Cerebellar ataxia	0	0	30
RF	Normal	24	1	5
	Sensory ataxia	0	30	0
	Cerebellar ataxia	0	2	28

BP Neural Network and SVM algorithm are above 97%. From **Table 4**, the source of error rate of RF algorithm is mainly used to identify the normal person between sensory ataxia and cerebellar ataxia; for gait recognition, BP Neural Network and SVM are considered; for Romberg's test, all three classification algorithms are suitable.

From **Table 3**, we can see that in Romberg's test, very high precision is achieved, and the error rate is mainly due to the misjudgment of normal and sensory ataxia. The reason for this is that in Romberg's test, cerebellar ataxia subjects are not stable whether they open or close their eyes, while sensory ataxia subjects and normal subjects remained stable during eye opening; the only difference between the two is that sensory ataxia subjects shake after closing their eyes, and closing eyes have no effect on normal people. The time domain waveform from **Figure 9** can also give the corresponding conclusion. The ability to maintain body balance is related to the age, gender and the length of time for standing of the individual. Normal people may have slight shaking in the Romberg's test, symptoms of patients with sensory ataxia may be mild, which may cause confusion. It is worth mentioning that it's difficult to distinguish normal subjects from sensory ataxia subjects completely in Romberg's test, but cerebellar ataxia can be detected. In order to ensure the experimental results more reliable, gait detection

experiments are also performed, and the two experiments confirmed reliability of the system.

CONCLUSION

Sensory Ataxia and Cerebellar Ataxia are neurological diseases which affect the patients' quality of life seriously; therefore, their detection at early stage are very important and necessary. In this paper, non-contact wireless sensing technology has been proposed to discriminate symptoms between the two diseases. The advantages include improvement of comfort, overcoming self-consciousness enhancing, etc. The main merit of the system lies in its convenience and price cost advantage. We firstly preprocess the data by removing outliers, wavelet transform filtering, then data features are extracted, finally, we use BP Neural Network, SVM, RF machine learning algorithms to train the model. The experimental results show that most of the algorithms can achieve more than 96% prediction accuracy, which can effectively discriminate between sensory ataxia and cerebellar ataxia, and prove that the technical scheme described in this paper is effective. Next, we will further explore the application of C-Band wireless sensing technology in healthcare, and propose more clinical application programs to make clinical detection more accurate, reliable and smarter, so as to reduce the burden on clinicians and patients.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Northwest Women's and Children's Hospital, Xi'an Jiaotong University Health Science Center. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

QZ: manuscript writing. XZ, YL, and XY: guidance. XY: editing, project management, and funding. QA: verification. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We would like to thank Mr. Guan, Mr. Fan, and Mr. Zhu for their help and guidance in the measurement and analysis.

REFERENCES

Afendi, T. M., Kurugollu, F., and Crookes, D. (2013). "Gait period estimation algorithm based on the angles of extremities," in *Proceedings of the European Workshop on Visual Information Processing (EUVIP), Paris, France* (Piscataway, NJ: IEEE).

Bastian, A. J. (1997). Mechanisms of ataxia. *Phys. Ther.* 77, 672–675.

Bastian, A. J., Martin, T. A., Keating, J. G., and Thach, W. T. (1996). Cerebellar ataxia: abnormal control of interaction torques across multiple joints. *J. Neurophysiol.* 76, 492–509. doi: 10.1152/jn.1996.76.1.492

Cazzato, D., Bella, E. D., Dacci, P., Mariotti, C., and Lauria, G. (2016). Cerebellar ataxia, neuropathy, and vestibular areflexia syndrome: a slowly progressive disorder with stereotypical presentation. *J. Neurol.* 263, 245–249. doi: 10.1007/s00415-015-7951-9

Chhetri, S., Gow, D., Shaunak, S., and Varma, A. (2014). Clinical assessment of the sensory ataxias; diagnostic algorithm with illustrative cases. *Prac. Neurol.* 14, 242–251. doi: 10.1136/practneurol-2013-000764

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

Demsar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* 7, 1–30. doi: 10.1155/2013/395096

Diener, H. C., and Dichgans, J. (1992). Pathophysiology of cerebellar ataxia. *Mov. Disord.* 7, 95–109. doi: 10.1002/mds.870070202

Diener, H. C., Dichgans, J., Bacher, M., and Gompf, B. (1984). Quantification of postural sway in normals and patients with cerebellar diseases. *Electroencephalogr. Clin. Neurophysiol.* 57, 134–142. doi: 10.1016/0013-4694(84)90172-x

Donnelly, K. (2011). *Sensory Ataxia*. New York, NY: Springer.

Fadic, R., Russell, J. A., Vedanarayanan, V. V., Lehar, M., Kuncl, R. W., and Johns, D. R. (1997). Sensory ataxic neuropathy as the presenting feature of a novel mitochondrial disease. *Neurology* 49, 239–245. doi: 10.1212/wnl.49.1.239

Franchignoni, F. P., Vanni, G., Savoini, C., Grioni, G., and Galante, M. (1984). Evaluation of sensory ataxia utilizing the computerized romberg test in peripheral neuropathies. *Riv. Neurobiol.* 30, 438–446.

Hain, T. C., and Cherchi, M. (2017). *Approach to the Patient with Dizziness and Vertigo Practical Neurology*, 5th Edn. London: Wolters Kluwer Health Pharma Solutions (Europe) Ltd, 207–226.

Havasi, L., Szlávik, Z., and Szirányi, T. (2007). Detection of gait characteristics for scene registration in video surveillance system. *IEEE Trans. Image Process.* 16, 503–510. doi: 10.1109/tip.2006.888339

Klockgether, T., and Paulson, H. (2011). Milestones in ataxia. *Mov. Disord.* 26, 1134–1141. doi: 10.1002/mds.23559

Lanska, D. J. (2002). The romberg sign and early instrument for measuring postural sway. *Semin. Neurol.* 22, 409–418. doi: 10.1055/s-2002-36763

Li, L., Wen, Z., and Wang, Z. (2016). *Outlier Detection and Correction During the Process of Groundwater Lever Monitoring Base on Pauta Criterion with Self-learning and Smooth Processing*. Berlin: Springer.

Lorincz, J., and Begusic, D. (2006). "Physical layer analysis of emerging IEEE 802.11n WLAN standard," in *Proceedings of the Advanced Communication Technology, Icact the International Conference* (Piscataway, NJ: IEEE).

Maurer, C., Mergner, T., Bolha, B., and Hlavacka, F. (2001). Human balance control during cutaneous stimulation of the plantar soles. *Neurosci. Lett.* 302, 45–48. doi: 10.1016/s0304-3940(01)01655-x

Mcgough, E. L., Hsu, L. Y., Thompson, H. J., and Teri, L. (2018). Concurrent validity of postural sway measures in older adults with cognitive impairment. *Phys. Occup. Ther. Geriatr.* 36, 399–410. doi: 10.1080/02703181.2018.1556231

Missaoui, B., Bendaya, S., Mane, M., and Thoumie, P. (2013). Balance and gait parameters in sensory ataxia; effects of a balance training program. *Ann. Phys. Rehabil. Med.* 56, e198–e207.

Mochizuki, H., and Ugawa, Y. (2010). Cerebellar ataxic gait. *Brain Nerve* 62, 1203–1210.

Pearce, J. M. S. (2005). Romberg and his sign. *Eur. Neurol.* 53, 210–213. doi: 10.1159/000086732

Poornachandra, S. (2008). Wavelet-based denoising using subband dependent threshold for ECG signals. *Digit. Signal Process.* 18, 49–55. doi: 10.1016/j.dsp.2007.09.006

Rumelhart, D. E., Durbin, R., Golden, R., and Chauvin, Y. (1995). "Backpropagation: the basic theory," in *Developments in Connectionist Theory. Backpropagation: Theory, Architectures, and Applications*, eds Y.

Chauvin, and D. E. Rumelhart (Mahwah, NJ: Lawrence Erlbaum Associates, Inc), 1–34.

Schmitz-Hübsch, T. (2006). Scale for the assessment and rating of ataxia (SARA). *Encycl.Mov. Disord*. 66, 95–99. doi: 10.1016/b978-0-12-374105-9.00534-7

Shi, T., and Horvath, S. (2006). Unsupervised learning with random forest predictors. *J. Comput. Graph. Stat*. 15, 118–138. doi: 10.1198/106186006x94072

Thomann, K. H., and Dul, M. W. (1996). Abnormal gait in neurologic disease. *Optom. Clin*. 5, 181–192.

Trouillas, P., Takayanagi, T., Hallett, M., Currier, R. D., Subramony, S. H., Wessel, K., et al. (1997). International cooperative ataxia rating scale for pharmacological assessment of the cerebellar syndrome. *J. Neurol. Sci*. 145, 205–211. doi: 10.1016/s0022-510x(96)00231-6

Umair Bin Altaf, M., Butko, T., and Juang, B. H. (2015). Acoustic gaits: gait analysis with footstep sounds. *IEEE Trans. Biomed. Eng*. 62, 2001–2011. doi: 10.1109/tbme.2015.2410142

Wang, X. C., Ristic-Durrant, D., Spranger, M., and Gräser, A. (2017). "Gait assessment system based on novel gait variability measures," in *Proceedings of the 2017 International Conference on Rehabilitation Robotics (ICORR), London, UK* (Piscataway, NJ: IEEE).

Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. *Chemometr. Intell. Lab. Syst*. 2, 37–52.

Yang, X., Shah, S. A., Ren, A., Nan, Z., and Jie, T. (2017). Wandering pattern sensing at S-band. *IEEE J. Biomed. Health Inform*. 22, 1863–1870. doi: 10.1109/jbhi.2017.2787595

Zheng, Y., Liu, Y., and Zhou, Z. (2013). From RSSI to CSI: indoor localization via channel response. *ACM Comput. Surv*. 46, 1–32. doi: 10.1145/2543581.2543592

Zhu, M., and Zhang, H. (2010). "Research on model of indoor distance measurement based on receiving signal strength," in *Proceedings of the International Conference on Computer Design & Applications* (Piscataway, NJ: IEEE).

Zongyi, L., and Sarkar, S. (2006). Improved gait recognition by gait dynamics normalization. *IEEE Trans. Pattern Anal. Mach. Intell*. 28, 863–876. doi: 10.1109/tpami.2006.122

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

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership