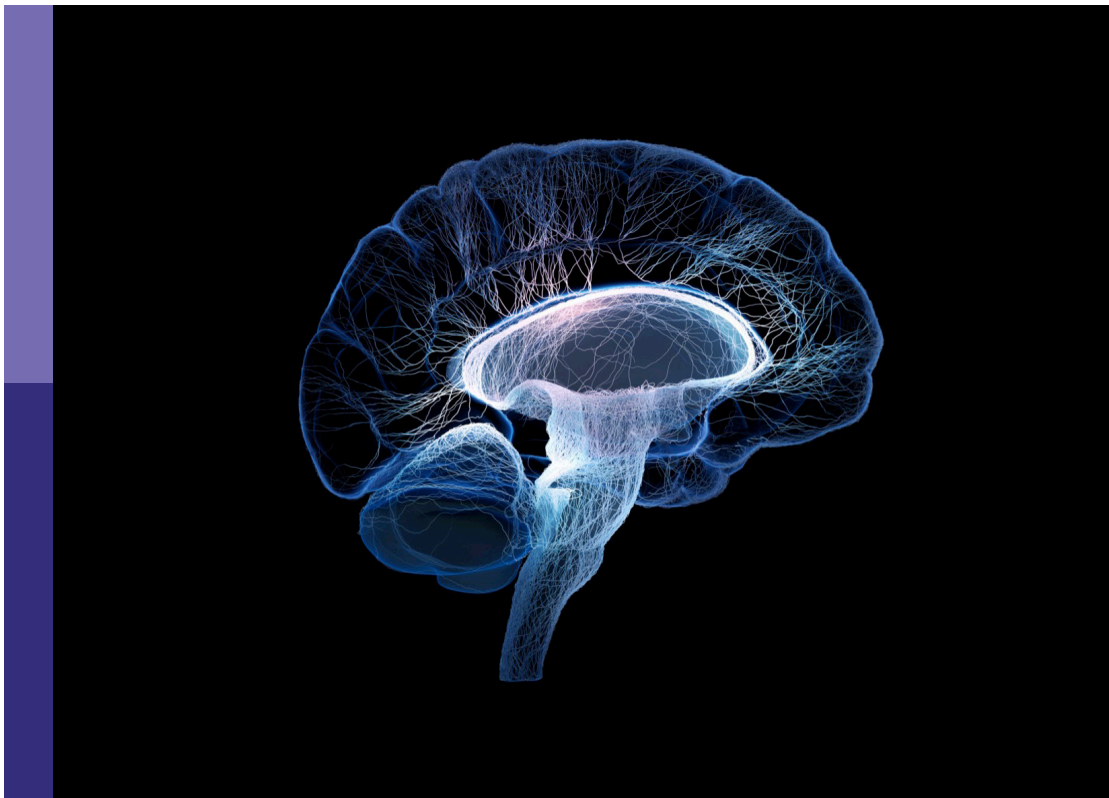
frontiers | Research Topics

Ear-centered sensing: From sensing principles to research and clinical devices, volume II

Edited by
Martin Georg Bleichner, Preben Kidmose and Jérémie Voix

Published in
Frontiers in Neuroscience



ISSN 1664-8714
ISBN 978-2-8325-4054-1
DOI 10.3389/978-2-8325-4054-1

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

Ear-centered sensing: From sensing principles to research and clinical devices, volume II

Topic editors

Martin Georg Bleichner — University of Oldenburg, Germany

Preben Kidmose — Aarhus University, Denmark

Jérémie Voix — École de technologie supérieure (ÉTS), Canada

Citation

Bleichner, M. G., Kidmose, P., Voix, J., eds. (2023). *Ear-centered sensing: From sensing principles to research and clinical devices, volume II*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-4054-1

Table of contents

frontiers | Frontiers in Neuroscience

TYPE Editorial
PUBLISHED 17 November 2023
DOI 10.3389/fnins.2023.1327801

OPEN ACCESS

EDITED AND REVIEWED BY
Laura Ballerini,
International School for Advanced Studies (SISSA), Italy

*CORRESPONDENCE
Jérémie Voix
✉ jeremie.voix@etsmtl.ca

RECEIVED 25 October 2023
ACCEPTED 26 October 2023
PUBLISHED 17 November 2023

CITATION
Voix J, Kidmose P and Bleichner MG (2023) Editorial: Ear-centered sensing: from sensing principles to research and clinical devices, volume II. *Front. Neurosci.* 17:1327801. doi: 10.3389/fnins.2023.1327801

Editorial: Ear-centered sensing: from sensing principles to research and clinical devices, volume II

Jérémie Voix[1]*, Preben Kidmose[2] and Martin Georg Bleichner[3]

[1]Université du Québec (ÉTS), Montreal, QC, Canada, [2]Department of Electrical and Computer Engineering, Aarhus University, Aarhus, Denmark, [3]Department of Psychology, University of Oldenburg, Oldenburg, Lower Saxony, Germany

KEYWORDS

EEG, brain activity, in-ear, brain-computer interface (BCI), wearable and mobile computing

Editorial on the Research Topic
Ear-centered sensing: from sensing principles to research and clinical devices, volume II

Welcome to the second volume of the "*Ear-centered sensing: from sensing principles to research and clinical devices*" Research Topic. In this Research Topic of nine articles, we delve deeper into the realm of ear-centered sensing, exploring its multifaceted applications and shedding light on the immense potential this field holds. Our aim, as Topic Editors, is to provide you with a ome of latest research and developments within this emerging field.

Ear-centered sensing: "the ear beyond hearing"

Ear-centered sensing, or "the ear beyond hearing" as coined by one of the Topic Editors (Voix, 2017), represents a groundbreaking approach to monitoring physiological signals, offering an unobtrusive and convenient means of data acquisition in everyday life. Physiological parameters such as heart rate, respiratory rate, eye blink and motion signals, skin conductance, as well as electrical activity from muscles and the brain, can all be captured from the ear. This unique positioning not only allows for discreet monitoring but also supports extended data recording, fostering a deeper understanding of psycho-physiological processes.

Ear-centered sensing is poised to play a pivotal role in scientific, diagnostic, and therapeutic endeavors, with a particular emphasis on mobile health applications. The challenges we face in this domain are intriguing, as we seek to bridge the gap between unconventional monitoring locations and established recording sites. Questions regarding signal fidelity, sensitivity, and real-time artifact discrimination remain at the forefront of our exploration.

Advancements in sensor technology

Dedicated sensor and amplifier technology are paramount for successful long-term usage of ear-centered sensing devices. Authors that contributed their papers in this Research Topic

have made significant strides in this area, striving for unobtrusiveness in every aspect. From biocompatible materials that adapt to individual anatomy to lightweight, inconspicuous instrumentation, their innovations promise to enable seamless data acquisition without restricting users' daily activities.

The path forward: contributions to ear-centered sensing

In this volume II of our Research Topic, we present a diverse range of research articles that collectively advance the field of ear-centered sensing. We have compiled a brief overview of each article to pique your interest:

1. "*Ear-EEG measures of auditory attention to continuous speech*": This study explores the potential of ear-centered EEG to monitor auditory attention during continuous speech, offering insights into assistive devices for complex auditory environments (Holtze et al.).

2. "*Evaluation of real-time endogenous brain-computer interface developed using ear-electroencephalography*": Investigating the feasibility of real-time endogenous ear-EEG-based brain-computer interfaces, this article explores the potential of ear-EEG in online environments (Choi et al.).

3. "*In-ear electro-oculography for attended speaker estimation*": Utilizing in-ear electro-oculography, this research focuses on improving comprehension in hearing-impaired individuals during conversations, with implications for hearing assistive devices (Skoglund et al.).

4. "*Pre-gelled electrode grid for self-applied EEG sleep monitoring at home*": Addressing the need for convenient home sleep monitoring, this article presents a self-applicable EEG sensor array for accurate sleep evaluation (da Silva Souto et al.).

5. "*Assessing focus through ear-EEG*": Investigating the potential of ear-EEG to determine levels of attention and focus, this study explores the integration of ear-EEG into wearable devices for monitoring mental load (Crétot-Richert et al.).

6. "*Synchronization of ear-EEG and audio streams in a portable research hearing device*": This article assesses the alignment of audio and EEG data in the context of hearing aid algorithms, offering insights into future closed-loop EEG and audio applications (Dasenbrock et al.).

7. "*Sound localization in children with unilateral microtia and atresia*": Exploring sound localization in children with hearing conditions, this study highlights the benefits of non-surgical bone conduction devices (Liu et al.).

8. "*At-home sleep monitoring using generic ear-EEG*": Introducing a generic ear-EEG device for at-home sleep monitoring, this research emphasizes the potential for widespread sleep stage monitoring (Tabar et al.).

9. "*Ear-EEG sensitivity modeling for neural sources and ocular artifacts*": This study establishes the sensitivity of ear-EEG to neural sources and ocular artifacts, supporting its integration into EEG paradigms (Yarici et al.).

As you delve into these nine articles, we hope you gain a deeper appreciation for the transformative potential of ear-centered sensing across various domains, from healthcare to cognitive science. By bringing together experts from different disciplines, we aim to foster collaboration and innovation in this every-day growing field.

Author contributions

JV: Writing—original draft, Writing—review & editing. PK: Writing—review & editing. MB: Writing—review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

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

Publisher's note

References

Voix, J. (2017). "The ear beyond hearing: from smart earplug to in-ear brain computer interfaces," in *The 24th International Congress on Sound and Vibration (ICSV24)*. London: International Institute of Acoustics and Vibration (IIAV).

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 24 March 2022
doi: 10.3389/fnins.2022.842635

Evaluation of Real-Time Endogenous Brain-Computer Interface Developed Using Ear-Electroencephalography

*Soo-In Choi[1], Ji-Yoon Lee[2,3], Ki Moo Lim[1,4] and Han-Jeong Hwang[2,3]**

[1] Department of Medical IT Convergence Engineering, Kumoh National Institute of Technology, Gumi-si, South Korea, [2] Department of Electronics and Information Engineering, Korea University, Sejong City, South Korea, [3] Interdisciplinary Graduate Program for Artificial Intelligence Smart Convergence Technology, Korea University, Sejong City, South Korea, [4] Department of IT Convergence Engineering, Kumoh National Institute of Technology, Gumi-si, South Korea

While previous studies have demonstrated the feasibility of using ear-electroencephalography (ear-EEG) for the development of brain-computer interfaces (BCIs), most of them have been performed using exogenous paradigms in offline environments. To verify the reliable feasibility of constructing ear-EEG-based BCIs, the feasibility of using ear-EEG should be further demonstrated using another BCI paradigm, namely the endogenous paradigm, in real-time online environments. Exogenous and endogenous BCIs are to use the EEG evoked by external stimuli and induced by self-modulation, respectively. In this study, we investigated whether an endogenous ear-EEG-based BCI with reasonable performance can be implemented in online environments that mimic real-world scenarios. To this end, we used three different mental tasks, i.e., mental arithmetic, word association, and mental singing, and performed BCI experiments with fourteen subjects on three different days to investigate not only the reliability of a real-time endogenous ear-EEG-based BCI, but also its test-retest reliability. The mean online classification accuracy was almost 70%, which was equivalent to a marginal accuracy for a practical two-class BCI (70%), demonstrating the feasibility of using ear-EEG for the development of real-time endogenous BCIs, but further studies should follow to improve its performance enough to be used for practical ear-EEG-based BCI applications.

Keywords: electroencephalography (EEG), ear-EEG, brain-computer interface (BCI), endogenous BCI, test-retest reliability

OPEN ACCESS

Edited by:
Jérémie Voix,
École de Technologie Supérieure (ÉTS), Canada

Reviewed by:
Dan Zhang,
Tsinghua University, China
Vassiliy Tsytsarev,
University of Maryland, College Park, United States

***Correspondence:**
Han-Jeong Hwang
hwanghj@korea.ac.kr

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

Received: *23 December 2021*
Accepted: *03 March 2022*
Published: *24 March 2022*

Citation:
Choi S-I, Lee J-Y, Lim KM and Hwang H-J (2022) Evaluation of Real-Time Endogenous Brain-Computer Interface Developed Using Ear-Electroencephalography. Front. Neurosci. 16:842635. doi: 10.3389/fnins.2022.842635

INTRODUCTION

A brain-computer interface (BCI) provides a potential alternative to the normal communication method, which involves languages and body movements, for disabled patients such as those with locked-in syndrome and aphasia (Bauer et al., 1979; Gao et al., 2021; Xu et al., 2021). It translates neuronal brain activity measured invasively or non-invasively into commands for controlling external devices, such as wheelchairs, robot arms, and computers (Hwang et al., 2013a).

Most BCIs have been realized using non-invasive neuroimaging modalities for measuring brain activity on the scalp, such as electroencephalography (EEG), magnetoencephalography (Mellinger et al., 2007), and near-infrared spectroscopy (Power et al., 2012). Among the modalities, EEG has

been the most widely used owing to its reasonable cost, portability, and high temporal resolution (Hwang et al., 2013a). However, BCIs based on the traditional scalp-EEG have several disadvantages from the viewpoint of their usability (Looney et al., 2012; Debener et al., 2015; Goverdovsky et al., 2016). They require measurement electrodes to be attached to the scalp with conductive gels, which is time-consuming. Moreover, the electrode attachment time increases with the number of electrodes to be mounted on the scalp. For instance, it generally exceeds 0.5 h for 30 electrodes, which leads to both the subject and operator being exhausted during the preparation for EEG. Furthermore, the electrodes mounted on the scalp are unaesthetic, and subjects need to wash their hair to remove the conductive gels after the BCI experiment. These disadvantages of conventional scalp-EEG-based BCIs render BCI technology difficult to be used outside laboratory environments. As an alternative to the traditional scalp-EEG used in BCIs, some researchers have proposed the use of EEG in which brain activity is measured around or inside the ears; such EEG is termed ear-EEG (Looney et al., 2012; Debener et al., 2015; Bleichner and Debener, 2017; Choi et al., 2018; Kaveh et al., 2020), and measurements can be obtained with a miniaturized and compact hardware system.

The feasibility of using ear-EEG in BCIs has been verified in many previous studies (Bleichner et al., 2015; Bleichner et al., 2016; Fiedler et al., 2016; Fiedler et al., 2017; Choi et al., 2018; Floriano et al., 2018; Wei et al., 2018). Most of the previous studies on ear-EEG-based BCIs have used exogenous paradigms involving external auditory or visual stimuli to evoke stimuli-specific brain activity, such as auditory steady-state response (ASSR) (Kidmose et al., 2012; Looney et al., 2012; Kidmose et al., 2013a,b; Mikkelsen et al., 2015; Goverdovsky et al., 2016; Bech Christensen et al., 2017; Goverdovsky et al., 2017), steady-state visual evoked potential (SSVEP) (Norton et al., 2015; Goverdovsky et al., 2016; Goverdovsky et al., 2017), and event-related potential (ERP) (Kidmose et al., 2012; Bleichner et al., 2015; Debener et al., 2015; Norton et al., 2015; Fiedler et al., 2016; Pacharra et al., 2017). In our previous study (Choi et al., 2018), we showed that an ear-EEG-based BCI can be developed using an endogenous paradigm involving self-modulated brain activity, without any external stimuli, and we classified two different mental states induced during mental arithmetic (MA) and resting state with an average accuracy of about 78%. We also proposed an optimal re-referencing method to improve the signal-to-noise ratio (SNR) of ear-EEG and used it to improve the performance of an endogenous ear-EEG-based BCI (Choi and Hwang, 2019).

Although previous studies have shown the feasibility of using ear-EEG in BCIs, the practical usability of ear-EEG-based BCIs should be verified *via* real-time online experiments that mimic real-life scenarios. To the best of our knowledge, only two studies have introduced online ear-EEG-based BCIs by using a representative exogenous paradigm, SSVEP (Norton et al., 2015; Wang et al., 2015). However, no study has verified the feasibility of an endogenous BCI based on ear-EEG in online environments. To further demonstrate the potential possibility of using ear-EEG on the development of a practically usable BCI, the feasibility of using ear-EEG in developing endogenous BCIs should be also demonstrated, particularly in online experimental environments.

Accordingly, in the present study, we investigated whether an endogenous ear-EEG-based BCI could be reliably implemented with reasonable performance in real-world applications. We developed an online endogenous ear-EEG-based BCI and tested it on two days to investigate not only its reliability as a real-time endogenous ear-EEG-based BCI, but also its test-retest reliability. In the experiment, three mental tasks were employed: MA, mental singing (MS), and word association (WA). An offline experiment was first conducted to determine the best pair of mental tasks for each subject, and online experiments were then conducted on 2 days using individually selected best pairs of mental tasks.

MATERIALS AND METHODS

Subjects

Fourteen healthy subjects (average age: 25.57 $\pm$ 1.70 years; eight males and six females) participated in this study. All of them had normal or corrected-to-normal vision and hearing. None of them reported previous neurological, psychiatric, or other related diseases that could have affected the outcomes of this study. To minimize the impact of the subjects' physical condition on the experiment, the subjects were asked to sleep for at least 6 h on the day preceding the experiment and to avoid alcohol intake for at least 24 h before the experiment. Information regarding the detailed procedure of this study was provided before the experiment. All subjects gave informed consent prior to the beginning of the experiments and received monetary compensation after the experiment. This study was approved by the Institutional Review Board of the Kumoh National Institute of Technology (No. 6250), and was performed in accordance with relevant guidelines and regulations.

Ear-Electroencephalography (EEG) Measurement

The experiment was conducted in a soundproof room, and the subjects were seated in comfortable armchairs in front of a 21-in. monitor (LG, 24MP58VQ, Seoul, South Korea) and binaural speakers (Britz, BR-1000A, Cuve Black 2, Paju, South Korea). Ear-EEG data were recorded using eight electrodes attached behind the ears (four electrodes for each ear), as shown in **Figure 1A**, at a sampling rate of 1,000 Hz (actiCHamp, Brain Products GmbH Ltd., Gilching, Germany). In accordance with the international 10–20 system, the reference and ground electrodes were attached at the FCz and Fpz positions, respectively, but the reference effect was removed by re-referencing with only ear-EEG (**Figure 1B**), which is explained later in this paper. The impedance was maintained below 10 kΩ throughout the experiment.

Experimental Paradigm

Offline Experiment Conducted on Day 1

The objective of the offline experiment was to select the best pair of mental tasks for each subject, which was used in the subsequent

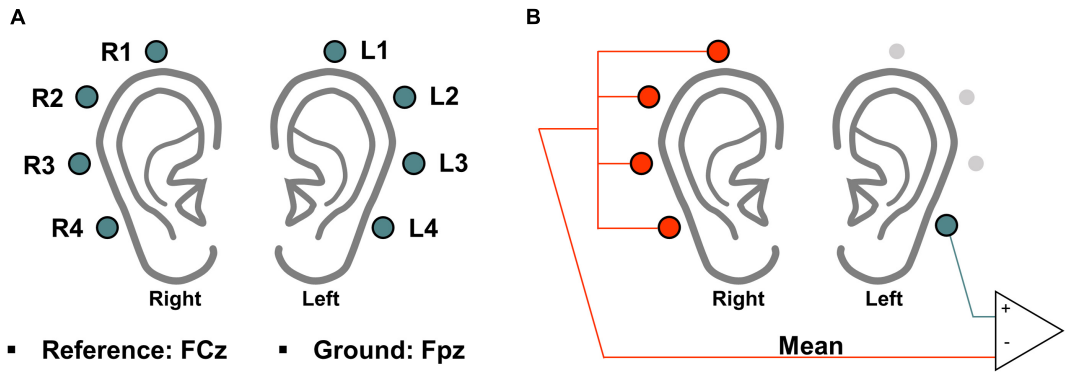


FIGURE 1 | (A) Schematic of electrode positions used to record ear-EEG data. **(B)** Schematic sketch of the re-referencing method used to remove the impact of the original reference electrode (FCz). Re-referencing was performed by subtracting the mean value of the opposite ear's electrodes from the value of an electrode of interest for every time point.

online experiments conducted on two different days. The offline experiment was hence performed by considering the following three mental tasks:

(i) MA: Continuously subtracting a single-digit number (between 5 and 9) from a three-digit number (e.g., 477 − 8 = 469, 469 − 8 = 461, . . .), with both numbers being randomly presented (Shin et al., 2016; Choi et al., 2018).

(ii) WA: Generating words beginning with a letter provided to the subjects in their native language (Korean; e.g., Apple, Arrow, Aerospace, . . . for "A") (Friedrich et al., 2012, 2013). The subjects had to generate as many words as they could.

(iii) MS: Mentally singing the English alphabet song from A to Z at a constant speed of 1 Hz, which induced relatively lower cognitive load compared with MA and WA (Shin et al., 2016).

Figure 2A shows the experimental paradigm of the offline experiment. Each subject completed five experimental sessions. At the beginning of each session, a blank image was presented for 5 s, during which the subjects were instructed to relax and prepare to act according to the upcoming instruction. After the rest period, eyes-closed (EC) and eyes-open (EO) tasks were sequentially performed, each for 15 s, with the aim of verifying the reliability of ear-EEG measurements based on alpha activity changes (Looney et al., 2012; Goverdovsky et al., 2016; Choi et al., 2018; Choi and Hwang, 2019). Subsequently, the three mental tasks were randomly performed, with each task being performed 10 times, in a single session. A single trial comprised task presentation for 5 s, subsequent task execution for 10 s, and a variable rest period that ranged from 8 to 13 s. During the task presentation period, a combination of three-digit and one-digit numbers for MA (e.g., 477 – 8), a single letter for WA (e.g., "A"), or the string "ABC" for MS was randomly presented on the monitor. The corresponding mental task was performed for the following 10 s, during which time the subjects gazed at a fixation mark presented at the center of the monitor to minimize eye movements. During the variable rest period, the word "rest" and an asterisk were presented on the monitor, and the subjects were instructed to gaze at the asterisk with a blank mind to minimize eye movements. The three mental tasks were randomly presented in a session, and each session was ended with EC and EO tasks. Each subject completed five sessions, performing 50 trials for each of the three mental tasks. There was a break of 3 to 5 min between the sessions to allow the subjects to rest.

Online Experiment Conducted on Day 2 and 3

On the basis of the classification accuracies of all combinations of the three mental tasks performed in the offline experiment (i.e., MA vs. MS, MA vs. WA, and MS vs. WA), we selected the best pair of mental tasks for each subject and used them for the following online experiments performed on two different days. The online experiment comprised three training and two test sessions, where we collected training data (30 trials for each task) using an individually selected best pair of mental tasks to construct an online classifier and we tested the classifier for new inputs (test data: 20 trials for each task) in real-time, respectively. The experimental paradigm of the online experiments was identical to that of the offline experiment, except that the best pair of mental tasks was employed for each subject and real-time feedback was provided in the two test sessions immediately following the task execution period, on the basis of the classification results (**Figure 2B**). In the first online experiment conducted on day 2, three training sessions were first performed. No feedback was provided during mental tasks in these sessions, whereas auditory feedback was provided on the basis of real-time classification results in the two test sessions following the task execution period. The experimental paradigm of the online experiment conducted on day 3 was identical to that used for the first online experiment conducted on day 2, except that the two test sessions were repeated thrice independently using three different online classifiers. We tested these three classifiers that were constructed using different training data sets, namely, the data set obtained on day 2, that acquired on day 3, and the combination of these two data sets, to

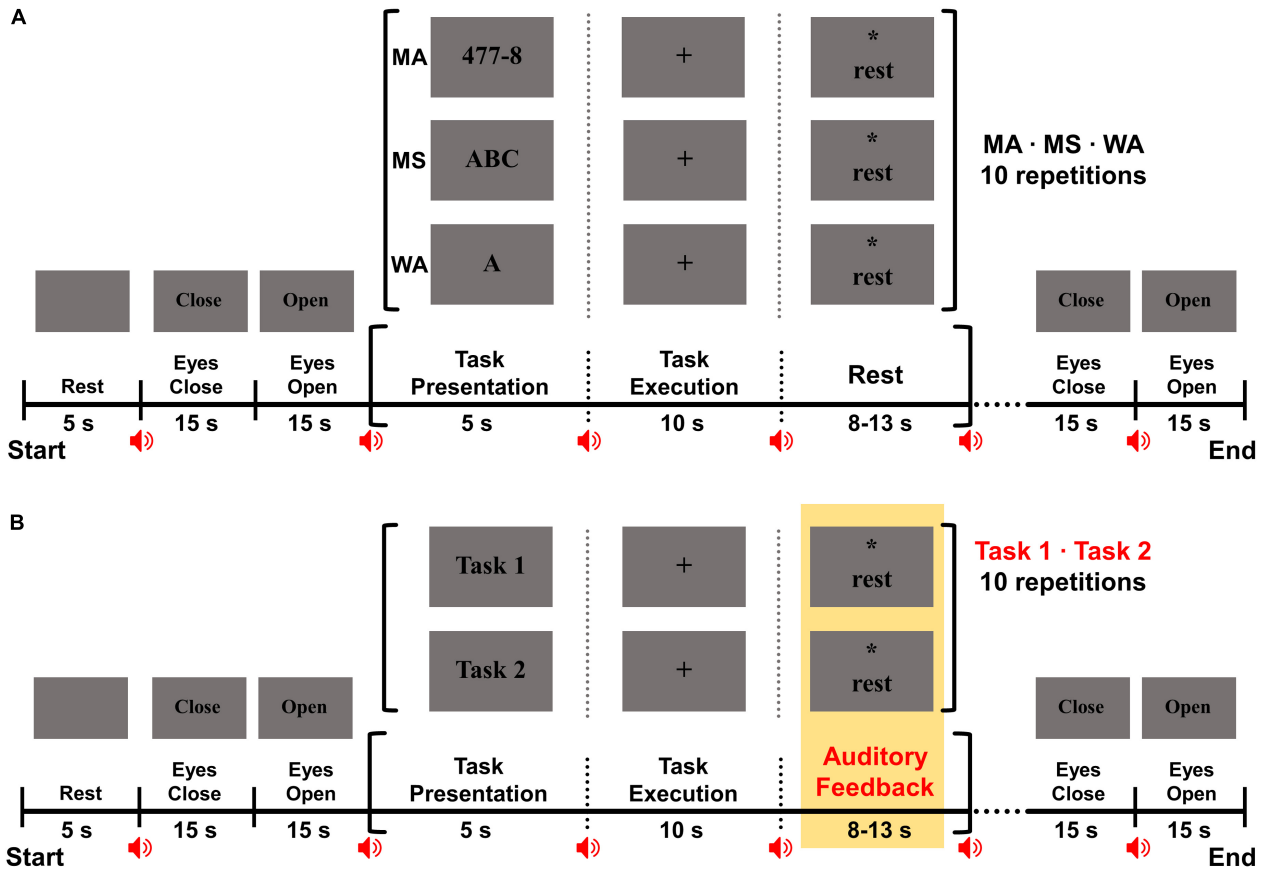


FIGURE 2 | (A) Schematic of the offline experimental paradigm. At the beginning of each session, a 5 s rest period was provided for preparing for the upcoming task. Before and after performing three mental tasks, the eyes were closed and opened for 15 s each to check the reliability of ear-EEG data related to alpha activity changes. A single trial comprised a 5 s task presentation period, a 10 s task execution period, and an 8 to 13 s rest period. During a task presentation period, a mental task that was to be performed was displayed for 5 s (e.g., 477 – 8 for MA, the string "ABC" for MS, and a Korean character for WA). During a task execution period, subjects were instructed to perform an indicated mental task for 10 s while focusing on a black fixation cross displayed at the center of a screen, to minimize eye movements. During a rest period, the word "rest" was presented along with an asterisk on the screen, and the subject was instructed to take a break without any thought and movement while staring at the asterisk. A short beep was presented at every screen transition (red speaker icons) to explicitly indicate the transition. **(B)** Schematic of the online experimental paradigm. The online experiment paradigm was identical to that of the offline experimental paradigm, except that an individually selected best pair of mental tasks was used for each subject and auditory feedback was provided on the basis of real-time classification results after task execution.

investigate the impact of different training data on their classification performance.

Ear-Electroencephalography (EEG) Data Analysis

Preprocessing

Data analysis was performed using MATLAB (MathWorks, Natick, MA, United States) along with the EEGLAB (Delorme and Makeig, 2004) and BBCI toolboxes (Blankertz et al., 2016), and the same analysis method was applied to the ear-EEG data measured in both offline and online experiments. Ear-EEG data were first band-pass filtered from 1 to 50 Hz using a zero-phase third-order Butterworth filter, and then down-sampled to 200 Hz to reduce the computation time. As the ear-EEG data were recorded using a reference at FCz, for the removal of the impact of the original reference electrode, all ear-EEG data were re-referenced using the mean value of the ear-EEG channels on the opposite ear because of the superior SNR of the re-referencing method (Choi and Hwang, 2019). Then, 10 s-epochs based on the task onset were extracted for each of the three mental tasks for classification. **Figure 1B** shows an example of the re-referencing method for one channel denoted by turquoise blue color on the left ear, and each of all channels were re-referenced by the same method.

Classification

A multiband common spatial pattern (CSP) was used to determine discriminative features for classification in five frequency bands (δ-band: 1–3 Hz, θ-band: 4–7 Hz, α-band: 8–13 Hz, β-band: 14–29 HZ, and γ-band: 30–50 Hz) (Ramoser et al., 2000; Lemm et al., 2005). A multiband CSP was independently applied to the 10-s epochs of the three mental tasks, respectively, for each of the five frequency bands, where the log-variances

TABLE 1 | Classification accuracies (%) of each pair of mental tasks for each subject in the offline experiment. The bold-font indicates the classification accuracies for the best pair of mental tasks for each subject.

	MA vs. MS	MA vs. WA	MS vs. WA	Best pair
Sub 1	**55.2**	55.1	42	55.2
Sub 2	76.7	**80.6**	64.7	80.6
Sub 3	**80.8**	65.3	70.6	80.8
Sub 4	63.7	70.2	**75.6**	75.6
Sub 5	51.4	67.3	**77.6**	77.6
Sub 6	**97.7**	96.8	93.6	97.7
Sub 7	55.8	54.2	**65.4**	65.4
Sub 8	**87.1**	48.4	83.5	87.1
Sub 9	**71.1**	65.8	58.7	71.1
Sub 10	81.4	**81.5**	72.2	81.5
Sub 11	**91.5**	83	79.1	91.5
Sub 12	74.9	**82.9**	67.5	82.9
Sub 13	**69.6**	51.2	61.7	69.6
Sub 14	66.8	**71.2**	60.7	71.2
Average	**73.1 ± 13.9**	**69.5 ± 14.2**	**69.5 ± 12.5**	**77.7 ± 10.9**

of the first and last CSP components were extracted for each frequency band as classification features. Shrinkage linear discriminant analysis (sLDA; Peck and Ness, 1982; Schäfer and Strimmer, 2005; Blankertz et al., 2011) was used as a classifier. Offline classification accuracies were obtained using 10-fold cross-validation for all possible pairs of the three mental tasks (MA vs. WA, MA vs. MS, and WA vs. MS) and they were used to select the best pair of mental tasks for each subject on day 1. Online classification accuracies were estimated using real-time outputs obtained immediately after performing a single trial in the two test sessions on days 2 and 3. As mentioned, two test sessions were performed thrice independently in the online experiment conducted on day 3. In the sessions, three different sLDA classifiers were constructed using different training data (data obtained on day 2, data acquired on day 3, and the combination of data obtained on these 2 days), and the online classification accuracies were independently estimated for each of the three classifiers.

Statistical Test

To investigate the feasibility of constructing an ear-EEG-based online BCI, we compared the online classification performance of our ear-EEG-based BCI with a theoretical 95% confidence limit of a chance accuracy on the basis of the number of trials for a two-class BCI (e.g., 59.61% for 50 trials of each task) (Müller-Putz et al., 2008). Furthermore, we employed a non-parametric statistical method, the Friedman test, for performing multiple comparisons with the Wilcoxon signed-rank test for *post hoc* in terms of the classification performance since the number of samples was insufficient (<30) for using parametric methods (i.e., RM-ANOVA) (Hesterberg, 2015).

Event-Related (de)Synchronization

To visually inspect task-specific brain activity in terms of event-related (de)synchronization (ERD/ERS), event-related spectral perturbation (ERSP) of each mental task was estimated for each epoch extracted from −2 to 10 s on the basis of the task onset for each channel and subject (baseline period: −2 to 0 s), and ERSPs were averaged over all subjects (Friedrich et al., 2012). To quantitatively investigate changes in ERSP values over three experimental days, we estimated ERSP values of three experimental days for five frequency bands (δ-band: 1–3 Hz, θ-band: 4–7 Hz, α-band: 8–13 Hz, β-band: 14–29 HZ, and γ-band: 30–50 Hz) for each of three mental tasks.

RESULTS

Offline Experimental Results

Table 1 shows the classification accuracies for each pair of the three mental tasks and the best pairs of mental tasks for each subject. The average classification accuracies of MA vs. MS, MA vs. WA, and MS vs. WA were 73.1 ± 13.9%, 69.5 ± 14.2%, and 69.5 ± 12.5%, respectively. The mean classification accuracy of the best pairs of mental tasks for each subject was 77.7 ± 10.9% (last column of **Table 1**), which was similar to that of our previous study (Choi and Hwang, 2019). Statistical test results showed that the mean classification accuracy of individually selected best pairs of mental tasks was significantly higher than those of MA vs. WA and MS vs. WA (Friedman: $\chi 2(3) = 15.88$, $p = 0.0012$, best pair > MA vs. WA = MS vs. WA), whereas the mean classification accuracy of MA vs. MS was not statistically different from the others. The combination of MA and MS was mostly selected as the best pair of mental tasks: MA vs. MS = 7, MA vs. WA = 4, and WA vs. MS = 3.

Online Experimental Results

Figure 3 shows the mean classification accuracies for the two online experimental days (days 2 and 3), along with the best mean offline classification accuracy shown in **Table 1** (day 1: 77.7 ± 10.9%). The mean online classification accuracies for days 2 and 3 were 69.1 ± 14.5% and 65.7 ± 12.7% (when only training data measured on day 3 was used), respectively, which were not statistically different from the best mean offline classification accuracy, despite the reduced performances. The mean online classification accuracy for day 3 decreased when the training data obtained on a different day (day 2) was used to build a classifier (61.9 ± 13.2%); by contrast, it increased when a combination of the training data acquired on two days (day 2 + day 3) was used (69.5 ± 14.7%). Despite the inter-experimental variability of the classification accuracies, all the mean online classification accuracies were higher than the theoretical 95% confidence limit of a chance accuracy (59.61%), and a classification accuracy of nearly 70% (marginal classification accuracy for a practical two-class BCI) was obtained for two online experimental days (for day 3, when a combination of training data obtained on days 2 and 3 was used).

Event-Related (de)Synchronization Maps

Figure 4 shows the grand-average ERD/ERS maps of all subjects for each mental task over the 3 days; the maps were obtained by averaging all channels. Overall, for all three mental tasks, on day 1 (offline experiment), widespread ERS was observed

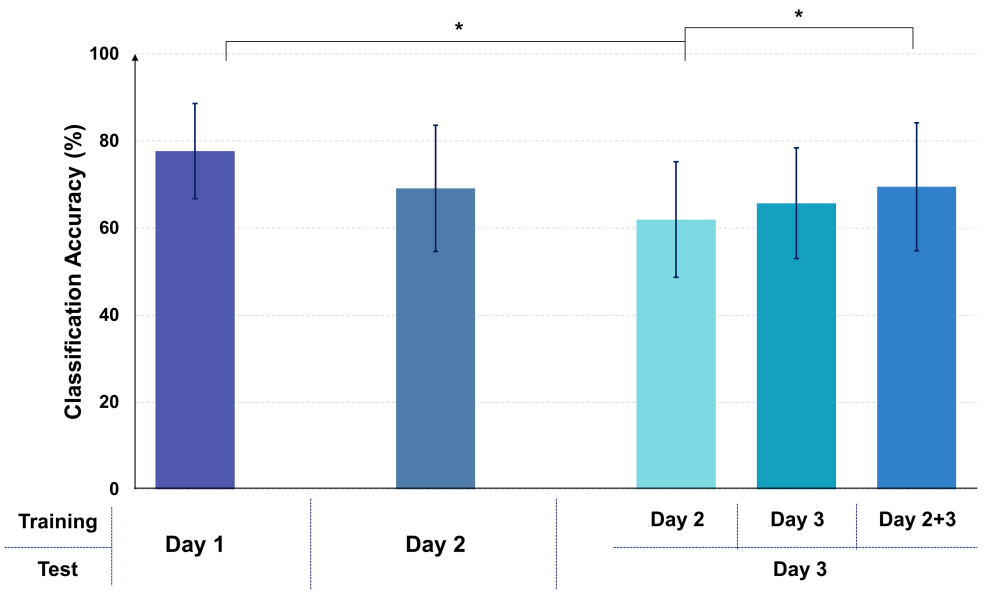


FIGURE 3 | Mean classification accuracies of the offline experiment (day 1) and the two online experiments (days 2 and 3). The mean offline classification accuracy was obtained by averaging the offline classification accuracies for the best pairs of mental tasks for each subject (day 1). The mean classification accuracies of day 3 were obtained using three different training data sets (day 2, day 3, and day 2 + day 3). An asterisk indicates a statistically significant difference.

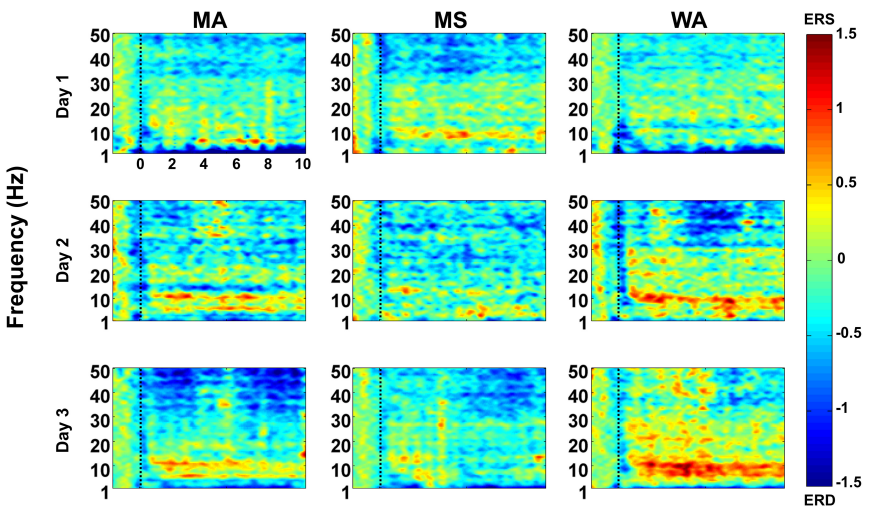


FIGURE 4 | ERD/ERS pattern maps averaged over all channels and subjects over the three experimental days. The x-axis indicates the task period between −2 to 10 s on the basis of the task onset, and the y-axis represents the frequency ranging from 1 to 50 Hz. The task onset at t = 0 s is marked by a dotted line.

in a middle frequency band (7–30 Hz), while relatively strong ERD was observed in the other frequency band (first row of **Figure 4**). However, despite the similar ERD/ERS patterns, each mental task showed relatively distinct ERD/ERS patterns, making it possible to distinguish each of the tasks, such as relatively stronger δ- and θ-ERD for MA than MS while stronger γ-ERD for MS than MA. Moreover, relatively small changes were observed in ERD/ERS for WA compared with MA and MS. The ERD/ERS pattern for MA— α- and β-ERS and γ-ERD— tended to become stronger from the first experimental day to the last experimental day (day 1 → day 2 → day 3 in the first column of **Figure 4**). Furthermore, WA particularly started to show its unique ERD/ERS pattern from the first online experiment (day 2), namely, widespread strong ERS over the entire frequency band along with strong γ-ERD. However, small changes in ERD/ERS were observed for MS over the three experimental days. The quantitative analysis results for ERD/ERS pattern changes over three experimental days are provided with respect to five frequency bands (δ-band: 1–3 Hz, θ-band: 4–7 Hz, α-band: 8–13 Hz, β-band: 14–29 Hz, and γ-band: 30–50 Hz) for each of three mental tasks in **Supplementary Figure 1**.

DISCUSSION

Recently, ear-EEG has gained considerable attention for its potential use in the development of BCIs owing to its high usability and portability for brain activity measurement, despite its relatively low performance compared with scalp-EEG-based BCIs (Pacharra et al., 2017; Denk et al., 2018; Kaongoen et al., 2021). Most of the previously proposed ear-EEG-based BCIs have involved exogenous paradigms, such as SSVEP (Norton et al., 2015; Goverdovsky et al., 2016; Goverdovsky et al., 2017), ASSR (Mikkelsen et al., 2015; Bech Christensen et al., 2017), and ERP (Looney et al., 2012; Debener et al., 2015; Fiedler et al., 2016; Pacharra et al., 2017), and their feasibility has been demonstrated *via* offline experiments in general. However, the feasibility of using ear-EEG in the development of endogenous BCIs should be also demonstrated, especially in online experimental settings that mimic real-world scenarios, in order to extend the feasibility of using ear-EEG in the development of BCIs. In this study, we investigated the feasibility of an ear-EEG-based BCI involving self-modulated EEG by using an endogenous paradigm in an online environment, and we demonstrated that a real-time endogenous BCI can be implemented using ear-EEG, despite its decreased classification performance compared with that in offline experiments.

The offline classification accuracy of each mental task pair differed among the subjects. The mean offline classification accuracy for the best pair of mental tasks (77.7 ± 10.9%) was significantly higher than those of the other two pairs of mental tasks (**Table 1**), namely MA vs. WA (69.5 ± 14.2%) and MS vs. WA (69.5 ± 12.5%), indicating the importance of using individualized mental tasks for the development of reliable BCIs (Hwang et al., 2014). The mean online classification accuracies for days 2 and 3 were lower by about 8 and 12% (day 2: 69.1 ± 14.5%; day 3: 65.7 ± 12.7%), respectively, compared with that of the offline experiment (77.7 ± 10.9%). The lower classification performance in the online experiment was probably because of the difference in the experimental environment between the training and test sessions; the difference depended on whether real-time feedback regarding classification results was provided (Shenoy et al., 2006). In general, the environmental difference between offline and online experiments gave rise to the inherent non-stationarity of EEG data, thereby resulting in a shift in the data distributions in feature space and ultimately degrading the BCI classification performance (Shenoy et al., 2006). On the other hand, real-time feedback provided in the online experiment reduced the impact of inter-day (session) variability on classification performance because the subject could adapt a given classifier by chaining a control strategy based on real-time feedback (Hwang et al., 2017); two classifiers trained using the EEG data measured on days 2 and 3, respectively, showed a small difference of classification accuracy when testing the EEG data measured on day 3 (61.9 ± 13.2% vs. 65.7 ± 12.7%). Despite the reduced classification performance in the online experiment, we obtained a meaningful mean classification accuracy of nearly 70%, which is a marginal classification accuracy for practical communication in two-class BCIs. Thus, we demonstrated that ear-EEG could be used to realize real-time endogenous BCIs. Nevertheless, the overall classification performance of our proposed ear-EEG-based BCI should be improved in order to increase the reliability of an ear-EEG-based endogenous BCI. Because the classification performance drop was partly compensated when the amount of training data was increased on day 3 (65.7 ± 12.7% → 69.5 ± 14.7% when a combination of training data of days 2 and 3 was used), it is expected that the classification accuracy will naturally increase and stabilize as a user uses a BCI over several days, owing to the cumulative data. Another approach to prevent the classification performance drop in the online experiment would be to introduce adaptive algorithms that use new data measured in real-time feedback sessions for classifier adaptation (Blankertz et al., 2007). We intend to work toward enhancing the performance of the proposed real-time ear-EEG-based BCI with the objective of developing reliable practical ear-EEG-based BCIs.

In this study, we employed three mental tasks (MA, MS, and WA) that have been widely used in previous BCI studies, and they showed somewhat overlapping but unique ERD/ERS patterns. MA showed relatively strong ERS in α- and β-bands and widespread ERD in a high frequency band (γ-band), which were similar to observations made in our previous studies (Choi et al., 2018; Choi and Hwang, 2019). Interestingly, the ERD/ERS pattern of MA became more dominant with the passage of time; from a neurophysiological viewpoint, this can be attributed to learning through real-time feedback in the online experiment, which led to better facilitation of brain activity (Duan et al., 2021). The ERD/ERS pattern of WA somewhat overlapped with that of MA in terms of α- and β-ERS with γ-ERD, but was also different from MA (e.g., relatively stronger θ- and α-ERS and shorter period of γ-ERD compared with MA). In particular, θ- and α-ERS became considerably stronger over the days for WA, which could be attributed to the learning, similar to MA. Unlike MA and WA, however, MS did not show significant changes in ERD/ERS patterns over the days, which could be explained by MS being the mental task originally designed to induce low cognitive load without intensively involving the brain resources during the mental task (Choi et al., 2018; Choi and Hwang, 2019). The qualitative results shown in **Figure 4** were also confirmed by the quantitative results shown in **Supplementary Figure 1**, such as an increasing trend of γ-ERD for MA and that of α-ERS for WA over three experimental days. Although the unique ERD/ERS patterns of each mental task tended to become stronger over the days because of learning through real-time feedback, as mentioned above, the overall classification performance rather decreased from day 1 to day 3, which might be because of the overlapping ERD/ERS patterns between the mental tasks despite their unique ERD/ERS patterns, such as the relatively strong ERS and ERD in the low- and high-frequency bands, respectively, as well as increased non-stationarity of EEG data between the training and test sessions in the online experiment. Therefore, it is necessary to develop more advanced algorithms that can fully utilize task-specific ERD/ERS patterns that change over the days, to improve the overall performance of endogenous ear-EEG-based BCIs.

Because BCI technology was introduced to help paralyzed patients to communicate with the outside world, BCI spellers have been widely developed, mainly using exogenous paradigms, such as SSVEP (Hwang et al., 2013b; Lim et al., 2015) and ERP (Gong et al., 2020; Miao et al., 2020). As an alternative to exogenous BCI spellers that require auditory or visual stimuli, imagined speech has been also actively studied for communication purposes. The auditory cortex close to the ears is responsible for speech and has shown task-specific brain activity during imagined speech (Kaongoen et al., 2021). In this study, we used three mental tasks that have been most widely used in BCI studies, but imagined speech might be the mental task most suitable for implementing ear-EEG-based BCIs since ear-EEG can capture task-specific brain activity generated from the auditory cortex more reliably compared with other brain areas owing to the adjacency effect. Therefore, it would be interesting to investigate the feasibility of using imagined speech in ear-EEG-based endogenous BCIs.

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 Institutional Review Board of the Kumoh National Institute of Technology. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

S-IC, KL, and H-JH designed the experiment and wrote the manuscript. S-IC and J-YL acquired the data. S-IC and H-JH performed the data analysis. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Basic Research Program through the National Research Foundation of Korea (NRF) funded by the MSIT (No. 2020R1A4A1017775), the Ministry of Education (No. 2019R1I1A3A01060732), and the South Korea Government (MSIP; Ministry of Science, ICT and Future Planning) (No. 2017R1C1B5017909). This research was also partly supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) and funded by the Korean Government (MSIT) (No. NRF-2021M3E5D2A01019547).

ACKNOWLEDGMENTS

We thank all subjects who participated in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.842635/full#supplementary-material

REFERENCES

Bauer, G., Gerstenbrand, F., and Rumpl, E. (1979). Varieties of the locked-in syndrome. *J. Neurol.* 221, 77–91. doi: 10.1007/BF00313105

Bech Christensen, C., Harte, J. M., Lunner, T., and Kidmose, P. (2017). Ear-EEG based objective hearing threshold estimation evaluated on normal hearing subjects. *IEEE Trans. Biomed. Eng.* 65, 1026–1034. doi: 10.1109/TBME.2017.2737700

Blankertz, B., Acqualagna, L., Dähne, S., Haufe, S., Schultze-Kraft, M., Sturm, I., et al. (2016). The Berlin brain-computer interface: progress beyond communication and control. *Front. Neurosci.* 10:530. doi: 10.3389/fnins.2016.00530

Blankertz, B., Dornhege, G., Krauledat, M., Müller, K.-R., and Curio, G. (2007). The non-invasive Berlin brain-computer Interface: fast acquisition of effective performance in untrained subjects. *Neuroimage* 37, 539–550. doi: 10.1016/j.neuroimage.2007.01.051

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

Bleichner, M. G., and Debener, S. (2017). Concealed, unobtrusive ear-centered EEG acquisition: cEEGrids for transparent EEG. *Front. Hum. Neurosci.* 11:163. doi: 10.3389/fnhum.2017.00163

Bleichner, M. G., Lundbeck, M., Selisky, M., Minow, F., Jager, M., Emkes, R., et al. (2015). Exploring miniaturized EEG electrodes for brain-computer interfaces. An EEG you do not see? *Physiol. Rep.* 3:e12362. doi: 10.14814/phy2.12362

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

Choi, S.-I., Han, C.-H., Choi, G.-Y., Shin, J., Song, K. S., Im, C.-H., et al. (2018). On the feasibility of using an ear-EEG to develop an endogenous brain-computer interface. *Sensors* 18:9. doi: 10.3390/s18092856

Choi, S.-I., and Hwang, H.-J. (2019). Effects of different re-referencing methods on spontaneously generated ear-EEG. *Front. Neurosci.* 13:822. doi: 10.3389/fnins.2019.00822

Debener, S., Emkes, R., De Vos, M., and Bleichner, M. (2015). Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear. *Sci. Rep.* 5:16743. doi: 10.1038/srep16743

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

Denk, F., Grzybowski, M., Ernst, S. M. A., Kollmeier, B., Debener, S., and Bleichner, M. G. (2018). Event-related potentials measured from in and around the ear electrodes integrated in a live hearing device for monitoring sound perception. *Trends Hear.* 22:2331216518788219. doi: 10.1177/2331216518788219

Duan, X., Xie, S., Xie, X., Obermayer, K., Cui, Y., and Wang, Z. (2021). An online data visualization feedback protocol for motor imagery-based BCI training. *Front. Hum. Neurosci.* 15:625983. doi: 10.3389/fnhum.2021.625983

Fiedler, L., Obleser, J., Lunner, T., and Graversen, C. (2016). "Ear-EEG allows extraction of neural responses in challenging listening scenarios - a future technology for hearing aids?," in *Proceedings of the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Orlando, FL, 5697–5700. doi: 10.1109/EMBC.2016.7592020

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

Floriano, A., Diez, P. F., and Freire Bastos-Filho, T. (2018). Evaluating the influence of chromatic and luminance stimuli on SSVEPs from behind-the-ears and occipital areas. *Sensors* 18:615. doi: 10.3390/s18020615

Friedrich, E. V., Scherer, R., and Neuper, C. (2012). The effect of distinct mental strategies on classification performance for brain-computer interfaces. *Int. J. Psychophysiol.* 84, 86–94. doi: 10.1016/j.ijpsycho.2012.01.014

Friedrich, E. V., Scherer, R., and Neuper, C. (2013). Stability of event-related (de-) synchronization during brain-computer interface-relevant mental tasks. *Clin. Neurophysiol.* 124, 61–69. doi: 10.1016/j.clinph.2012.05.020

Gao, X., Wang, Y., Chen, X., and Gao, S. (2021). Interface, interaction, and intelligence in generalized brain-computer interfaces. *Trends Cogn. Sci.* 25, 671–684. doi: 10.1016/j.tics.2021.04.003

Gong, M., Xu, G., Li, M., and Lin, F. (2020). An idle state-detecting method based on transient visual evoked potentials for an asynchronous ERP-based BCI. *J. Neurosci. Methods* 337:108670. doi: 10.1016/j.jneumeth.2020.108670

Goverdovsky, V., Looney, D., Kidmose, P., and Mandic, D. P. (2016). In-ear EEG from viscoelastic generic earpieces: robust and unobtrusive 24/7 monitoring. *IEEE Sens. J.* 16, 271–277. doi: 10.1109/JSEN.2015.2471183

Goverdovsky, V., Von Rosenberg, W., Nakamura, T., Looney, D., Sharp, D. J., Papavassiliou, C., et al. (2017). Hearables: multimodal physiological in-ear sensing. *Sci. Rep.* 7:6948. doi: 10.1038/s41598-017-06925-2

Hesterberg, T. C. (2015). What teachers should know about the bootstrap: resampling in the undergraduate statistics curriculum. *Am. Stat.* 69, 371–386. doi: 10.1080/00031305.2015.1089789

Hwang, H.-J., Kim, S., Choi, S., and Im, C.-H. (2013a). EEG-based brain-computer interfaces: a thorough literature survey. *Int. J. Hum. Comput. Interact.* 29, 814–826. doi: 10.1080/10447318.2013.780869

Hwang, H.-J., Kim, D. H., Han, C.-H., and Im, C.-H. (2013b). A new dual-frequency stimulation method to increase the number of visual stimuli for multi-class SSVEP-based brain-computer interface (BCI). *Brain Res.* 1515, 66–77. doi: 10.1016/j.brainres.2013.03.050

Hwang, H.-J., Lim, J.-H., Kim, D.-W., and Im, C.-H. (2014). Evaluation of various mental task combinations for near-infrared spectroscopy-based brain-computer interfaces. *J. Biomed. Opt.* 19:77005. doi: 10.1117/1.JBO.19.7.077005

Hwang, H.-J., Hahne, J. M., and Müller, K.-R. (2017). Real-time robustness evaluation of regression based myoelectric control against arm position change and donning/doffing. *PLoS One* 12:e0186318. doi: 10.1371/journal.pone.0186318

Kaongoen, N., Choi, J., and Jo, S. (2021). Speech-imagery-based brai-computer interface system using ear-EEG. *J. Neural Eng.* 18:016023. doi: 10.1088/1741-2552/abd10e

Kaveh, R., Doong, J., Zhou, A., Schwendeman, C., Gopalan, K., Burghardt, F. L., et al. (2020). Wireless user-generic ear EEG. *IEEE Trans. Biomed. Circuits Syst.* 14, 727–737. doi: 10.1109/TBCAS.2020.3001265

Kidmose, P., Looney, D., Jochumsen, L., and Mandic, D. P. (2013a). "Ear-EEG from generic earpieces: a feasibility study," in *Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Osaka, 543–546. doi: 10.1109/EMBC.2013.6609557

Kidmose, P., Looney, D., Ungstrup, M., Rank, M. L., and Mandic, D. P. (2013b). A study of evoked potentials from ear-EEG. *IEEE Trans. Biomed. Eng.* 60, 2824–2830. doi: 10.1109/TBME.2013.2264956

Kidmose, P., Looney, D., and Mandic, D. P. (2012). "Auditory evoked responses from ear-EEG recordings," in *Proceedings of the 2012 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, San Diego, CA, 586–589. doi: 10.1109/EMBC.2012.6345999

Lemm, S., Blankertz, B., Curio, G., and Muller, K.-R. (2005). Spatio-spectral filters for improving the classification of single trial EEG. *IEEE Trans. Biomed. Eng.* 52, 1541–1548. doi: 10.1109/TBME.2005.851521

Lim, J.-H., Lee, J.-H., Hwang, H.-J., Kim, D. H., and Im, C.-H. (2015). Development of a hybrid mental spelling system combining SSVEP-based brain–computer interface and webcam-based eye tracking. *Biomed. Signal Process. Control* 21, 99–104. doi: 10.1016/j.bspc.2015.05.012

Looney, D., Kidmose, P., Park, C., Ungstrup, M., Rank, M., Rosenkranz, K., et al. (2012). The in-the-ear recording concept: user-centered and wearable brain monitoring. *IEEE Pulse* 3, 32–42. doi: 10.1109/MPUL.2012.2216717

Mellinger, J., Schalk, G., Braun, C., Preissl, H., Rosenstiel, W., Birbaumer, N., et al. (2007). An MEG-based brain–computer interface (BCI). *NeuroImage* 36, 581–593. doi: 10.1016/j.neuroimage.2007.03.019

Miao, Y., Yin, E., Allison, B. Z., Zhang, Y., Chen, Y., Dong, Y., et al. (2020). An ERP-based BCI with peripheral stimuli: validation with ALS patients. *Cogn. Neurodyn.* 14, 21–33. doi: 10.1007/s11571-019-09541-0

Mikkelsen, K. B., Kappel, S. L., Mandic, D. P., and Kidmose, P. (2015). EEG recorded from the ear: characterizing the ear-EEG method. *Front. Neurosci.* 9:438. doi: 10.3389/fnins.2015.00438

Müller-Putz, G., Scherer, R., Brunner, C., Leeb, R., and Pfurtscheller, G. (2008). Better than random? A closer look on BCI results. *Int. J. Bioelectromagn.* 10, 52–55.

Norton, J. J. S., Lee, D. S., Lee, J. W., Lee, W., Kwon, O., Won, P., et al. (2015). Soft, curved electrode systems capable of integration on the auricle as a persistent brain–computer interface. *Proc. Natl. Acad. Sci. U.S.A.* 112, 3920–3925. doi: 10.1073/pnas.1424875112

Pacharra, M., Debener, S., and Wascher, E. (2017). Concealed around-the-ear EEG captures cognitive processing in a visual simon task. *Front. Hum. Neurosci.* 11:290. doi: 10.3389/fnhum.2017.00290

Peck, R., and Ness, J. V. (1982). The use of shrinkage estimators in linear discriminant analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 4, 530–537. doi: 10.1109/TPAMI.1982.4767298

Power, S. D., Kushki, A., and Chau, T. (2012). Automatic single-trial discrimination of mental arithmetic, mental singing and the no-control state from prefrontal activity: toward a three-state NIRS-BCI. *BMC Res. Notes* 5:141. doi: 10.1186/1756-0500-5-141

Ramoser, H., Muller-Gerking, J., and Pfurtscheller, G. (2000). Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans. Rehabil. Eng.* 8, 441–446. doi: 10.1109/86.895946

Schäfer, J., and Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat. Appl. Genet. Mol. Biol.* 4:32. doi: 10.2202/1544-6115.1175

Shenoy, P., Krauledat, M., Blankertz, B., Rao, R. P., and Müller, K.-R. (2006). Towards adaptive classification for BCI. *J Neural Eng.* 3, R13–R23. doi: 10.1088/1741-2560/3/1/R02

Shin, J., Müller, K.-R., and Hwang, H.-J. (2016). Near-infrared spectroscopy (NIRS)-based eyes-closed brain-computer interface (BCI) using prefrontal cortex activation due to mental arithmetic. *Sci. Rep.* 6:36203. doi: 10.1038/srep36203

Wang, Y., Nakanishi, M., Kappel, S. L., Kidmose, P., Mandic, D. P., Wang, Y., et al. (2015). "Developing an online steady-state visual evoked potential-based brain-computer interface system using earEEG," in *Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Milan, 2271–2274. doi: 10.1109/EMBC.2015.7318845

Wei, C. S., Wang, Y. T., Lin, C. T., and Jung, T. P. (2018). Toward drowsiness detection using non-hair-bearing EEG-based brain-computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 400–406. doi: 10.1109/TNSRE.2018.2790359

Xu, L., Xu, M., Jung, T. P., and Ming, D. (2021). Review of brain encoding and decoding mechanisms for EEG-based brain-computer interface. *Cogn. Neurodyn.* 15, 569–584. doi: 10.1007/s11571-021-09686-x

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

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 | Frontiers in Neuroscience

ORIGINAL RESEARCH
published: 03 May 2022
doi: 10.3389/fnins.2022.869426

Ear-EEG Measures of Auditory Attention to Continuous Speech

Björn Holtze[1], Marc Rosenkranz[2], Manuela Jaeger[1,3], Stefan Debener[1,4,5] and Bojana Mirkovic[1]*

[1] Neuropsychology Lab, Department of Psychology, University of Oldenburg, Oldenburg, Germany, [2] Neurophysiology of Everyday Life Group, Department of Psychology, University of Oldenburg, Oldenburg, Germany, [3] Division Hearing, Speech and Audio Technology, Fraunhofer Institute for Digital Media Technology IDMT, Oldenburg, Germany, [4] Research Center for Neurosensory Science, University of Oldenburg, Oldenburg, Germany, [5] Cluster of Excellence Hearing4all, University of Oldenburg, Oldenburg, Germany

Auditory attention is an important cognitive function used to separate relevant from irrelevant auditory information. However, most findings on attentional selection have been obtained in highly controlled laboratory settings using bulky recording setups and unnaturalistic stimuli. Recent advances in electroencephalography (EEG) facilitate the measurement of brain activity outside the laboratory, and around-the-ear sensors such as the cEEGrid promise unobtrusive acquisition. In parallel, methods such as speech envelope tracking, intersubject correlations and spectral entropy measures emerged which allow us to study attentional effects in the neural processing of natural, continuous auditory scenes. In the current study, we investigated whether these three attentional measures can be reliably obtained when using around-the-ear EEG. To this end, we analyzed the cEEGrid data of 36 participants who attended to one of two simultaneously presented speech streams. Speech envelope tracking results confirmed a reliable identification of the attended speaker from cEEGrid data. The accuracies in identifying the attended speaker increased when fitting the classification model to the individual. Artifact correction of the cEEGrid data with artifact subspace reconstruction did not increase the classification accuracy. Intersubject correlations were higher for those participants attending to the same speech stream than for those attending to different speech streams, replicating previously obtained results with high-density cap-EEG. We also found that spectral entropy decreased over time, possibly reflecting the decrease in the listener's level of attention. Overall, these results support the idea of using ear-EEG measurements to unobtrusively monitor auditory attention to continuous speech. This knowledge may help to develop assistive devices that support listeners separating relevant from irrelevant information in complex auditory environments.

Keywords: around-the-ear EEG, cEEGrid, auditory attention, speech envelope tracking, intersubject correlation (ISC), spectral entropy, auditory attention decoding (AAD)

OPEN ACCESS

Edited by:
Jérémie Voix,
École de Technologie Supérieure (ÉTS), Canada

Reviewed by:
Dan Zhang,
Tsinghua University, China
Philipp Ruhnau,
University of Central Lancashire, United Kingdom

****Correspondence:***
Björn Holtze
bjoern.holtze@uni-oldenburg.de

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

Received: *04 February 2022*
Accepted: *25 March 2022*
Published: *03 May 2022*

Citation:
Holtze B, Rosenkranz M, Jaeger M, Debener S and Mirkovic B (2022) Ear-EEG Measures of Auditory Attention to Continuous Speech. Front. Neurosci. 16:869426. doi: 10.3389/fnins.2022.869426

INTRODUCTION

In everyday complex auditory scenes, one fundamental question to be answered is how the brain manages to select relevant and neglect irrelevant information. Although many studies on auditory attention have contributed to this question, most of them have been conducted in highly controlled laboratory settings using discrete and artificial stimuli. Two recent advances have opened up the possibility of measuring brain responses to natural stimuli in everyday life. First, the development of small and portable measurement devices has made it possible to measure brain activity outside of the lab (e.g., Debener et al., 2012). Second, methods have been developed to analyze the neural processing of natural and continuous stimuli such as speech (Hamilton and Huth, 2020). Here, we investigate the potential of combining these two developments to eventually measure attentional processes unobtrusively.

Electroencephalography (EEG) is a popular method to non-invasively measure human brain electrical activity by placing electrodes on the scalp. Traditional EEG as used in most laboratories require caps or nets to position electrodes on the scalp, which is not feasible for EEG acquisition in everyday life (Bleichner and Debener, 2017). For unobtrusive EEG acquisition, small and near-invisible approaches are preferred to not disturb natural social interaction. This demand has led to the development of in-ear EEG (Looney et al., 2012), and around-the-ear EEG solutions (Debener et al., 2015), where electrodes are placed inside the outer ear canal or around the ear, respectively. The cEEGrid is one around-the-ear EEG solution – a c-shaped flex-printed sensor array comprising 10 electrodes (**Figure 1A**). In the current study, we used the cEEGrid as it provides larger inter-electrode distances compared to in-ear EEG, leading to an increase in the measured EEG amplitudes (Bleichner and Debener, 2017) and better sensitivity to distant contributions (Meiser et al., 2020).

Debener et al. (2015) and Bleichner et al. (2016) already provided evidence that attentional processes can be captured with cEEGrids but used event-based analyses and time-domain trial averaging instead of measuring the neural response to continuous stimuli. Here we investigated three methods which cannot only analyze the neural processing of continuous speech but have also been shown to be sensitive to attentional effects. We evaluated their feasibility to capture neural effects of auditory attention when using around-the-ear EEG.

The first method is speech envelope tracking, which refers to the neural tracking of the slow amplitude fluctuations, i.e., the envelope, of speech (Aiken and Picton, 2008). When presented with more than one speaker at the same time, the listener's neural signal correlates more strongly with the speech envelope of the attended than of the ignored speaker(s) (Power et al., 2012; Zion Golumbic et al., 2013). Based on this observation, many studies have been conducted to decode the attended among all present speakers from the listeners' EEG (Mirkovic et al., 2015; O'Sullivan et al., 2015). Therefore, this method is often referred to as auditory attention decoding (AAD). However, we will not use this term, because in principle all three methods introduced here aim at decoding the listener's attention. One of the main potentials of speech envelope tracking is that it can help to develop neuro-steered hearing aids that first identify and then enhance the attended speaker (Geirnaert et al., 2021b). This could have a tremendous impact for hearing-impaired listeners who have difficulties listening to one speaker in the presence of background noise (Shinn-Cunningham and Best, 2008).

The second method is known as intersubject correlations (ISCs). This method is based on the observations that individuals who are exposed to the same stimulus show similar spatiotemporal brain activity (Hasson et al., 2004; Dmochowski et al., 2012, for more recent reviews see Zhang, 2018 and Nastase et al., 2019). More recently, this approach has been adopted to attention research. When selectively attending to one of two simultaneously presented audio streams, ISCs of EEG signals were found to be higher for those participants attending to the same stream than for those attending to different streams (Stuldreher et al., 2020; Rosenkranz et al., 2021). Moreover, Rosenkranz et al. (2021) showed that the magnitude of participants' ISCs with others attending to the same audio stream were positively correlated with the participants' attentional effect observed in speech envelope tracking. Thus, the strength by which an individual's EEG signal correlates with others attending to the same stimulus reflects the individual's ability to selectively attend to the target stimulus and ignore the distracting stimulus. Regarding future application, this method could be of value in classroom scenarios (Poulsen et al., 2017; Janssen et al., 2021). For instance, this method could support students who have difficulties in focusing on the lecture content or support lecturers by identifying settings in which lectures are most effective (Brouwer et al., 2019).

The third method is spectral entropy. Spectral entropy characterizes the structure of an EEG spectrum (Viertiö-Oja et al., 2004) and has been proposed as a measure of attention. A high spectral entropy indicates an equally distributed EEG spectrum. This means that the power in each frequency band is very similar, whereas a low spectral entropy indicates an EEG spectrum in which the power is concentrated in one frequency band (Lesenfants and Francart, 2020). Lesenfants et al. (2018) found increased spectral entropy when participants actively attended to a stimulus compared to when they did not attend to the presented stimulus. In a consecutive study, Lesenfants and Francart (2020) showed that speech envelope tracking of a single speaker was increased during periods of high spectral entropy. Based on these findings, the authors concluded that high spectral entropy indicates high levels of attention. It is important to note that compared to speech envelope tracking and ISC, spectral entropy is not used to draw conclusion about one's direction of attention. Instead, spectral entropy may be an informative measure in scenarios where it is important to monitor one's level of attention, for example, when driving a car.

The aim of the current study was to test whether speech envelope tracking, ISCs, and spectral entropy capture effects of auditory attention to ongoing natural stimuli when unobtrusive around-the-ear EEG acquisition is used. To this end, we analyzed listeners' brain activity captured with cEEGrids while they attended to one of two simultaneously presented, continuous speech streams. As speech envelope tracking has previously been

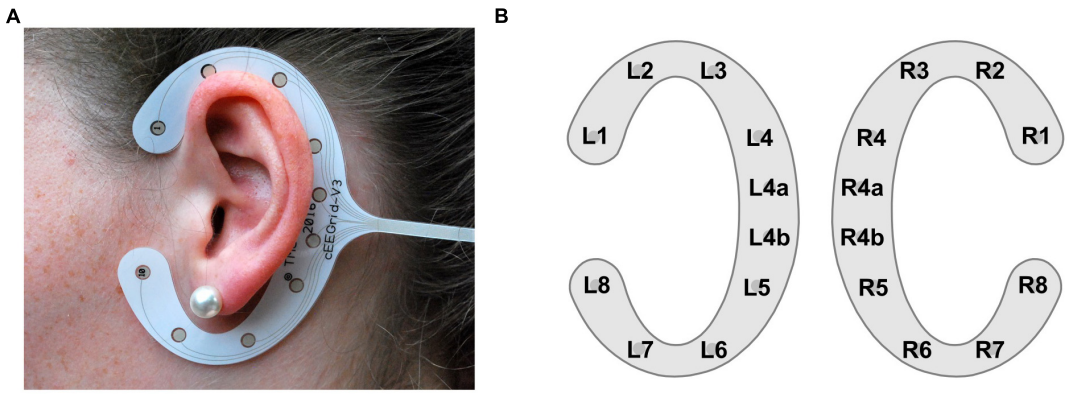


FIGURE 1 | cEEGrid illustration. **(A)** A cEEGrid attached with double-sided adhesive around the left ear. **(B)** cEEGrid channel layout consisting of a pair of cEEGrids, one for the left and one for the right ear. Each cEEGrid comprises 10 electrodes. Electrodes R4a and R4b serve as ground and analog reference, respectively. In the analysis, data were re-referenced to the half of channel L4b (algebraic linked mastoids). To keep the number of channels symmetrical between the left and right cEEGrid channel L4a was removed in the analysis.

performed on cEEGrid data but yielded rather low accuracies in identifying the attended speaker (Mirkovic et al., 2016; Nogueira et al., 2019), we explored the effect of artifact correction and of individualizing decoding models on the accuracy to identify the attended speaker.

MATERIALS AND METHODS

Participants

In the current study, two previously recorded, unpublished cEEGrid datasets were combined. The cEEGrid datasets were each concurrently recorded with cap EEG. The corresponding cap EEG datasets were originally used in Jaeger et al. (2020) and Holtze et al. (2021), respectively, and later jointly used in Rosenkranz et al. (2021). Here, we only considered the cEEGrid datasets. From Jaeger et al. (2020), five out of the 20 participants had to be excluded due to data loss during the cEEGrid recording. From Holtze et al. (2021) all 21 participants could be included, resulting in a total of 36 participants (mean age 23.6 years, 25 females) in the current study. All participants were native German speakers, had normal hearing based on audiometric thresholds of 20 dB HL or better in both ears at octave frequencies from 250 Hz to 8 kHz (Holmes and Griffiths, 2019), and reported no psychological or neurological condition. Both original studies were approved by the local ethics committee (University of Oldenburg, Germany, Jaeger et al., 2020, Drs.Nr.27/2018; Holtze et al., 2021, Drs.EK/2019/006). All participants signed written informed consent before taking part in the respective study and received monetary reimbursement afterward.

Task and Stimuli

Participants were comfortably seated in a dimly lit and sound-attenuated booth. They were instructed to attend to one of two simultaneously presented audio books and had to keep their attention on the same audio book throughout the experiment. To motivate participants and to make sure that they attended to the instructed audio book, participants had to answer questions related to the content of the to-be-attended audio book. Each audio book was narrated by a different male speaker (for further details see Mirkovic et al. (2016) where these stimuli were originally used). In Jaeger et al. (2020) each audio book was presented *via* a free-field loudspeaker located to the front-left (−45°) and front-right (+45°) side of the participant, respectively. In Holtze et al. (2021) the audio books were presented *via* earphones while the audio books were spatially separated at an angle of ±30° in azimuth using a head related transfer function (Kayser et al., 2009). Thus, in both studies, one audio book appeared to originate from the front left of the participant and the other one from the front right. The same audio books were used in both studies. The mode of presentation did not seem to affect the neural processing of the audio books as shown in Rosenkranz et al. (2021), where the cap-EEG data from Jaeger et al. (2020) and Holtze et al. (2021) were jointly analyzed. In both studies the audio books were presented in blocks of 10 min. In the Jaeger study the experiment consisted of six 10-min blocks while in the Holtze study it consisted of five 10-min blocks. Within the first 10-min block, both audio books were presented at equal volume in both studies. In the Jaeger study this was maintained for all remaining blocks. In the Holtze study only in two out of the remaining four blocks both audio books were presented at equal volume while in the other two blocks the to-be-attended audio book was enhanced. To keep the equal volume aspect constant across both studies, in the current study we only included the three blocks from the Holtze study where both audio books were presented equally loud. To keep the amount of data per participant constant across studies, we also selected only three blocks from the Jaeger study. This always included the first 10-min block plus two from the remaining blocks. Due to technical issues the cEEGrid data of some participants in the Jaeger study were not recorded during all blocks. For those participants where only three blocks were available, we used those. When more than three blocks were available, we pseudo-randomly selected two blocks such that blocks 2–5 were evenly represented across

participants. Thus, in the current study we used 30 min of data per participant, which always included the first 10-min block.

Data Acquisition

The cEEGrid recording procedure was identical for both original studies. For better electrode to skin conductance, the skin around the ears was prepared with abrasive gel (Abralyt HiCl, Easycap GmbH, Hersching, Germany) and cleaned with 70% alcohol. Thereafter, a small amount of abrasive gel was placed on the cEEGrid electrodes (TMSI, Oldenzaal, Netherlands; Debener et al., 2015) before it was attached with double-sided adhesive around the ear. Impedances were kept below 20 kΩ. Each participant was equipped with two cEEGrids, one around the left and one around the right ear. Electrodes R4a and R4b of the right cEEGrid served as ground and reference, respectively (**Figure 1B**). The two cEEGrids were connected to a 24-channel mobile amplifier (SMARTING, mBrainTrain, Belgrade, Serbia) which transmitted the data *via* Bluetooth to a recording computer. The cEEGrid data were acquired with a sampling rate of 500 Hz. The transmitted cEEGrid data as well as the onset markers of the 10-min blocks were integrated using the Lab Recorder software based on the Lab Streaming Layer[1] to time synchronize these data streams (Mullen et al., 2015).

Data Analysis

Preprocessing

All analysis steps were performed in MATLAB (R2019b, The Math-Works Inc., Natick, MA, United States), using custom scripts[2]. The cEEGrid data were processed with EEGLAB (version 2020.0; Delorme and Makeig, 2004) using the cEEGrid EEGLAB plugin[3] (version 0.9). To account for the constant delay between the onset markers of the 10-min blocks and the corresponding EEG, we presented 20 beep tones to the participant prior to the experiment. We then computed the grand average event-related potential (ERP) in response to these beep tones and shifted the cEEGrid data to align the N1 latency of the cEEGrid data to the N1 latency observed in the cap-EEG, which in turn had been corrected based on a timing test. This resulted in a time delay of 54 ms for the Jaeger study and 70 ms for the Holtze study. The difference can be explained by the different audio presentation setups as described above (for details see Jaeger et al., 2020 and Holtze et al., 2021). Then, the cEEGrid data were re-referenced to algebraically linked mastoids by re-referencing the data to the half of channel L4b. To keep the cEEGrid layout symmetrical between the left and right side, channel L4a was removed, leaving 16 cEEGrid channels per participant (**Figure 1B**) (Debener et al., 2015).

Artifact Correction

Artifact correction was performed using artifact subspace reconstruction (ASR; Mullen et al., 2015), implemented in the EEGLAB plugin clean_rawdata (version 2.4). ASR identifies and reconstructs segments containing artifacts based on the statistics of artifact-free calibration data. In the current study, no explicit calibration data was provided, instead the plugin function automatically selected artifact-free calibration data from the entire recording. The clean_rawdata wrapper function consists of multiple sub-functions. The sub-functions clean_flatlines and clean_channels were not used, to keep the number of channels constant for all participants, and because the interpolation of removed cEEGrid channels may not produce reliable results (cf. Kang et al., 2015). As ASR requires high-pass filtered data (Mullen et al., 2015), we used the clean_drift function within clean_rawdata with the default high-pass transition band from 0.25 to 0.75 Hz. As cutoff parameter for the clean_asr function we used a rather liberal value of 10, as cutoff values below 10 may be prone to remove brain data (Chang et al., 2018). The function clean_windows, which removes data segments that still contain artifacts after performing ASR, was not used as continuous signals were required for the analyzes.

Speech Envelope Tracking

As mentioned, speech envelope tracking has previously been implemented with cEEGrid data but yielded rather low accuracies when the aim was to identify the attended speaker (Mirkovic et al., 2016: 69.33% with 50 one-min-segments per participant; Nogueira et al., 2019: 59.79% with 48 one-min segments per participant). Therefore, we systematically investigated two adaptations of the analysis pipeline used in Mirkovic et al. (2016), with the goal of increasing the classification accuracy. For a better understanding we now first describe how we implemented the analysis pipeline described in Mirkovic et al. (2016) and then explain the adaptations.

To extract the attended and ignored speech envelopes the audio data were first normalized, by dividing them by their standard deviation. Then, the absolute Hilbert transform was computed, and low-pass filtered at 8 Hz (Butterworth, filter order: 3). Lastly, the filtered data were down-sampled to 64 Hz to reduce subsequent computation times. In accordance with the two speech envelopes, the cEEGrid data were also low pass filtered at a cutoff frequency of 8 Hz (finite impulse response filter, Hann windows, filter order: 100), and then high-pass filtered at a cutoff frequency of 2 Hz (finite impulse response filter, Hann windows, filter order: 500). Afterward, the filtered cEEGrid data were normalized by dividing them by their standard deviation, and then down-sampled to 64 Hz.

For speech envelope tracking, we implemented a decoding model, i.e., we trained a model on an individual's cEEGrid data to predict the attended speech envelope. For a better replicability, we implemented the decoding model within the mTRF toolbox (version 2.1; Crosse et al., 2016). For this, the individual's cEEGrid data and speech envelopes were first segmented into non-overlapping 60 s segments using the mTRFpartition function. This resulted in 30 segments of each speech envelope and the corresponding cEEGrid data for each participant. Using the function mTRFattncrossval, a decoder was trained on 29 segments of the attended speech envelope and the corresponding cEEGrid data. This decoder was then used to reconstruct the attended speech envelope of the left-out segment. Afterward, the

[1]https://github.com/labstreaminglayer

[2]https://doi.org/10.5281/zenodo.6379903

[3]https://doi.org/10.5281/zenodo.5946875

reconstructed speech envelope was correlated with the attended and ignored speech envelope of the left-out segment, respectively. The difference between these Pearson correlation coefficients ($Corr_{att}$–$Corr_{ign}$) is considered as the attentional gain. If the attentional gain was positive, the left-out segment was regarded as classified correctly. The prediction error was quantified as the mean squared error between the reconstructed and the actual speech envelope. The process of training a decoder on 29 segments and testing it on the left-out segment was repeated 30 times in a leave-one-out cross-validation manner (Stone, 1974). The decoding accuracy was then determined as the percentage of correctly classified segments. Chance level decoding accuracy was based on a binomial significance threshold.

In the decoding model, two important model hyperparameters require adjustment. One is the time lag window, which accounts for the time between the onset of the presented auditory stimulus and its cortical response. The other is the regularization parameter. Regularization is a technique to avoid overfitting and estimate reliable model parameters that generalize to unseen data (Holdgraf et al., 2017). Regularization is especially important in decoding models as it strongly affects the decoding accuracy (Wong et al., 2018). To closely follow the analysis pipeline used in Mirkovic et al. (2016), we applied Tikhonov regularization. Therefore, we estimated the optimal regularization parameter which we then multiplied with the regularization matrix (Crosse et al., 2016).

In line with Mirkovic et al. (2016), the optimal model hyperparameters, i.e., the time lag window and regularization parameter λ, were initially chosen on a group level. To this end, the grand average decoding accuracies were computed for different sets of hyperparameters. Potential time lag windows of 45 ms duration ranged from –115 to 620 ms, with 30 ms of overlap. Potential regularization parameters ranged from 10^{-5} to 10^{5} in factors of 10. As a result, for each participant we computed the decoding accuracy of 517 different sets of hyperparameters, based on 47 different time lag windows and 11 regularization parameters. We then selected the set of hyperparameters that yielded the largest grand average decoding accuracy. These group-level based hyperparameters were then used for all individual decoders. So far, we have described how we implemented the analysis pipeline as similarly used in Mirkovic et al. (2016). In the following two sections, we will explain the adaptations we made to explore the effect of artifact correction and individually chosen hyperparameters on the decoding accuracy.

Effect of Artifact Correction

The first adaptation was to include artifact correction into the analysis pipeline. As mentioned in Mirkovic et al. (2016), one possible reason for the low decoding accuracies was that no eye-, muscle- or movement-related artifacts were corrected for. Therefore, we investigated the effect of artifact correction on the decoding accuracy. To this end, we once performed artifact correction before the data were pass-band filtered between 2 and 8 Hz and compared it to the uncorrected data filtered between 2 and 8 Hz. To evaluate the impact of artifact correction, we compared the individual decoding accuracies between uncorrected and ASR-corrected data using a Wilcoxon signed rank test. To quantify how much data was modified by ASR and to what extent, we split the filtered data into consecutive 1-s segments and calculated the spectral power in the frequency range from 2 to 8 Hz. For each 1-s segment, we then averaged the spectral power over all channels and calculated the change in dB from uncorrected to ASR-corrected data.

Effect of Individually Chosen Hyperparameters

As a second adaptation we used individually chosen hyperparameters, instead of using group-level chosen hyperparameters. Specifically, individualizing the time lag window may help to increase decoding accuracies. As mentioned above, the time lag window accounts for the time between the stimulus onset and its cortical response. It is well known that cortical response lags vary across individuals (Lauter and Karzon, 1990), as can also be seen in Mirkovic et al. (2019). To the best of our knowledge, the effect of individualizing the regularization parameter for speech envelope tracking has not been investigated. Therefore, we also explored this adaptation. To select the optimal hyperparameters for each participant separately, we chose the set of hyperparameters which yielded the highest decoding accuracy for the individual. If multiple sets of hyperparameters fulfilled this criterion, we selected the one set among them which yielded the lowest prediction error. We then compared the individual decoding accuracies between the models using group-level and individually chosen hyperparameters with a Wilcoxon signed rank test.

Leave-one-out cross-validation (standard cross-validation) is a technique to train and test a model, such that the data which is used to train the model is different from the data which is used to test the model (Stone, 1974). Leave-one-out cross-validation is commonly applied in auditory attention decoding research (O'Sullivan et al., 2015). To compare our results to other studies, we also followed this approach when comparing the decoding accuracies of models using group-level or individually chosen hyperparameters. However, one aspect of this approach, which is sometimes neglected, is that when a model includes hyperparameters to be tuned, validating the model should be done on a yet another part of the data (Holdgraf et al., 2017). This procedure ensures that the selected hyperparameters do not only lead to high decoding accuracies on the data used to select them, but also on independent data. To account for this potential bias, we performed an additional analysis using nested cross-validation (Varma and Simon, 2006; Parvandeh et al., 2020). To this end, we first randomly selected 10 out of the 30 segments per participant for later validation of our model. The remaining 20 segments were then used in a leave-one-out cross-validation manner to find the optimal set of hyperparameters. Afterward, all these 20 segments were used to train the model with the selected set of hyperparameters. Finally, the model was validated by computing the decoding accuracy of the 10 initially left-out segments. This entire process was repeated 50 times so that at each iteration 10 different segments were randomly selected for later validation. In the end, the decoding accuracies were averaged over all 50 iterations. To test whether the results obtained in the initial analysis without independent

validation data were biased, we performed the nested cross-validation approach once with group-level chosen and once with individually chosen hyperparameters. The difference between the resulting decoding accuracies was statistically evaluated using a Wilcoxon signed rank test.

Intersubject Correlations

Cap-EEG-based ISCs within a two competing speaker paradigm have previously been analyzed by Rosenkranz et al. (2021). In the current study, we performed the ISC analysis on the simultaneously acquired cEEGrid data. The aim was to test whether the attentional effect of ISCs, as observed by Rosenkranz et al. (2021), can also be observed with unobtrusive around-the-ear EEG recordings. Therefore, we closely followed the analysis pipeline presented in Rosenkranz et al. (2021), which was largely based on the publicly available code[4] from Cohen and Parra (2016). As mentioned above, ISCs are based on the observation, that people who are exposed to the same stimulus show similar brain activity. In the current study, the first 10-min block was the only one which was included for all participants. Therefore, to leverage the statistical power of the entire sample size, for the ISC analysis we only used the first 10-min block. This was also done to closely follow the analysis performed in Rosenkranz et al. (2021).

To compute ISCs, the preprocessed cEEGrid data from the first 10-min block were first cleaned from artifacts as described above. After artifact correction the data were low-pass filtered at a cutoff frequency of 40 Hz (finite impulse response filter, Hann windows, filter order: 100), and then high-pass filtered at a cutoff frequency of 1 Hz (finite impulse response filter, Hann windows, filter order: 500). Lastly, the data were down-sampled to 250 Hz. Simply correlating the individual EEG channels between participants would not reveal a good estimate of the ISCs due to the low signal to noise ratio of EEG (Dmochowski et al., 2012). Therefore, Dmochowski et al. (2012) developed the correlated component analysis, which is available in the publicly available code from Cohen and Parra (2016). In the correlated component analysis, EEG channels are linearly projected such that the resulting components are maximally correlated between participants. Importantly, the number of resulting components is identical to the number of initial EEG channels. Lastly, the ISC scores of the three most correlating components were summed, resulting in a single ISC sum score per individual. Other components were neglected as their correlations have been shown to be close to chance (Ki et al., 2016).

Attentional Effect on Intersubject Correlations

When presented with two concurrent auditory streams, those participants attending to the same stream show higher ISC sum scores than those attending to different streams, even though all individuals are exposed to the same physical stimulus (Stuldreher et al., 2020; Rosenkranz et al., 2021). Here, we investigated whether this attentional effect could also be observed with cEEGrids. To test this, the cEEGrid data of each participant were correlated once with the cEEGrid data of all participants attending to the same audio book (ISC_{same}) and once with the cEEGrid data of all participants attending to the other audio book (ISC_{other}). Importantly, the projection vector of the correlated component analysis was computed on all but the to-be-correlated participant to reduce the risk of overfitting. For each participant, this resulted in 16 components (number of available cEEGrid channels) for the ISC_{same} condition and 16 components for the ISC_{other} condition. Within a participant, we then compared the ISC scores of the different components between the ISC_{same} and ISC_{other} condition. This we did for the three most correlating components individually as well as for their sum, using paired sample *t*-tests. The difference between an individual's ISC sum score in the same and other condition is considered as the attentional effect (ISC_{same}–ISC_{other}). To compute the chance level for ISC scores, we created chance-distributions with circular time-shifted data (Parra et al., 2018). For each participant the data were shifted to a different extent but all EEG channels within a participant were shifted equally. This disturbed the temporal alignment between the participants' EEG but kept the temporal and spatial structure within a participant unchanged. The process of randomly shifting the data and computing ISC scores for both conditions was repeated 100 times. This resulted in a distribution of ISC scores for each component and condition separately. The 95th percentiles of these distributions served as chance level.

In addition, we also classified whether a person attended to the left or right story based on their ISC scores (Rosenkranz et al., 2021). Therefore, we once computed the ISC scores of each participant with all participants attending to the left audio book (ISC_{left}) and once with all participants attending to the right audiobook (ISC_{right}). Thus, ISC_{left} and ISC_{right} reflect the synchrony of one participant with others attending to the left or right audio book, respectively. For this analysis we used two projection vectors, one was computed on participants who attended to the left story, and one was computed on participants who attended to the right story. Again, the to-be-correlated participant was left out when computing the projection vectors. Lastly, we summed the ISC scores of the three most correlating components and classified the direction of attention based on the ISC sum scores. Classification accuracy was calculated using the area under the receiver operator curve. Chance level accuracy was estimated by randomly assigning the class labels left and right and then calculating the corresponding area under the receiver operator curve (Ki et al., 2016). This was repeated 1000 times, each time randomizing the class labels. The 95th percentile of this distribution was then considered as chance level. Lastly, we evaluated the neurophysiological plausibility of the ISC components. For that, we computed the projection vectors of the correlated component analysis once for those participants attending to the left story, once for those attending to the right story and once for all participants. As the projection vectors are not directly physiologically interpretable, the projection vectors (spatial filters) were transformed into spatial patterns (Haufe et al., 2014).

Spectral Entropy

To compute the spectral entropy, the preprocessed cEEGrid data were first cleaned from artifacts as described above. Additionally, the function clean_channels was used to identify channels which correlated less than 0.6 with their robust estimate. These

[4]https://www.parralab.org/isc/

channels were later neglected when calculating the spectral entropy averaged over channels. Here, only for three participants one artifactual channel was identified. Spectral entropy was computed based on the analysis described in Lesenfants and Francart (2020). For each one-min segment and channel, the spectrum from 8 to 32 Hz was computed using multitaper spectral analysis (7 tapers, MATLAB function: pmtm). Each spectrum was then normalized by dividing each frequency power by the sum of all frequency powers in the range from 8 to 32 Hz (Viertiö-Oja et al., 2004). Thereby, the power of each individual spectrum was equalized to one which enabled the comparison between participants and channels. The spectral entropy was then computed as the product between the normalized frequency power of each frequency bin and the logarithm of its inverse. These were then summed over all frequency bins and normalized by one over the logarithm of the number of frequency bins. This resulted in spectral entropies ranging between zero and one. A spectral entropy value close to one reflects a spectrum in which the power of each frequency bin is similar, whereas a lower spectral entropy indicates a spectrum in which the power of the spectrum is concentrated in a few frequency bins. In the end, we had 30 (segments) times 16 (channels) spectral entropy values per participant. We did not have any prior assumptions on which cEEGrid channels to use. Therefore, we averaged the spectral entropy values over non-artifactual channels. Spectral entropy has been linked to the level of sustained attention, with higher values reflecting higher levels of attention (Lesenfants et al., 2018). There is both behavioral and neurophysiological evidence that the level of auditory attention decreases over time (Moore et al., 2017). Therefore, we investigated the spectral entropy over time by computing the Spearman rank correlation coefficient between the segment number and the corresponding spectral entropy. This we did for each participant individually as well as for the grand average spectral entropy. Alpha power (8–12 Hz) has also consistently been associated with attention (Foxe and Snyder, 2011; Klimesch, 2012). As it comprises one important frequency band when computing the spectral entropy based on the frequency spectrum from 8 to 32 Hz, we also investigated alpha power over time. For that we averaged the normalized frequency power from 8 to 12 Hz over all non-artifactual channels.

Relation Between Attentional Measures

To investigate the relation between the attentional gain in speech envelope tracking ($Corr_{att}$–$Corr_{ign}$) and the attentional effect of ISC sum scores (ISC_{same}–ISC_{other}), the time resolved attentional gain values in speech envelope were averaged over time. This resulted in one speech envelope gain value and one ISC sum difference score per participant. To investigate the relation between the attentional gain in speech envelope tracking ($Corr_{att}$–$Corr_{ign}$) and the spectral entropy we performed two analyses. Unlike Lesenfants and Francart (2020), we followed a correlation-based approach. In the first analysis, we correlated the time resolved speech envelope gain values with the time resolved spectral entropy values for each participant separately. In the second analysis, we first averaged the time resolved speech envelope gain and spectral entropy values over time to have one value pair per participant. We then correlated these value pairs for all participants. In all of the above-mentioned correlational analyses, the attentional gain of speech envelope tracking was computed using artifact corrected data and individual hyperparameters identified with standard cross-validation. Lastly, we correlated the attentional effect observed in the ISC sum scores (ISC_{same}–ISC_{other}) with the spectral entropy values averaged over time. To statistically evaluate the correlations, we performed Spearman rank correlations.

RESULTS

Speech Envelope Tracking

When performing speech envelope tracking without artifact correction, the grand average decoding accuracies reached 71.3% (**Figure 2A**). Removing artifacts with ASR resulted in a grand average decoding accuracy of 72.13%, which was, however, not significantly higher (**Figure 2A**, Wilcoxon signed rank test, $Z = 0.84$, $p = 0.4$). In this analysis, the group-level chosen time lag window from 95 to 140 ms and a regularization parameter of 10^{-2} were used (**Figure 2B**, black rectangle). Most part of the data was not strongly modified by artifact correction. In fact, the change in spectral power (8–12 Hz) due to ASR was less than ±0.1 dB in 73.41% of all 1-s segments. In only 6.12% of all 1-s segments, the spectral power was changed more than ±3 dB (**Supplementary Figure 1**). Even though artifact correction did not significantly increase the decoding accuracy, all further analyses were performed on artifact corrected data to ensure that decoding the attended speaker is based on brain data and not on artifacts. Using individually chosen hyperparameters instead of group level chosen ones significantly increased the decoding accuracies to 82.59% (**Figure 2C**, Wilcoxon signed rank test, $Z = 5.04$, $p < 0.001$). The individually chosen optimal hyperparameters are shown in **Figure 2B**. However, when further controlling for overfitting with nested cross-validation the decoding accuracies dropped substantially and the group-level chosen hyperparameters outperformed those of individually chosen hyperparameters (**Figure 2D**, Wilcoxon signed rank test, $Z = -3.88$, $p < 0.001$).

Intersubject Correlations

Using cEEGrid data, we confirmed the expected effect that ISC scores were significantly higher for participants attending to the same audio book than for those attending to different audio books (**Figures 3A,C**). This was the case for the ISC sum scores, i.e., the sum of ISC scores of the three strongest components (**Figure 3A**, paired sample *t*-test, $t = 8.24$, $p < 0.001$), as well as for the ISC scores of the first two components (**Figure 3C**, paired sample *t*-test, component 1: $t = 7.93$, $p < 0.001$, component 2: $t = 6.2$, $p < 0.001$). For the third component there was no evidence for a difference in ISC scores between the same and other conditions (**Figure 3C**, paired sample *t*-test, $t = 1.5$, $p = 0.14$). Only the ISC score of the first and second component revealed above chance level effects. The ISC sum scores of each individual participant with all those attending to the left and right audio book

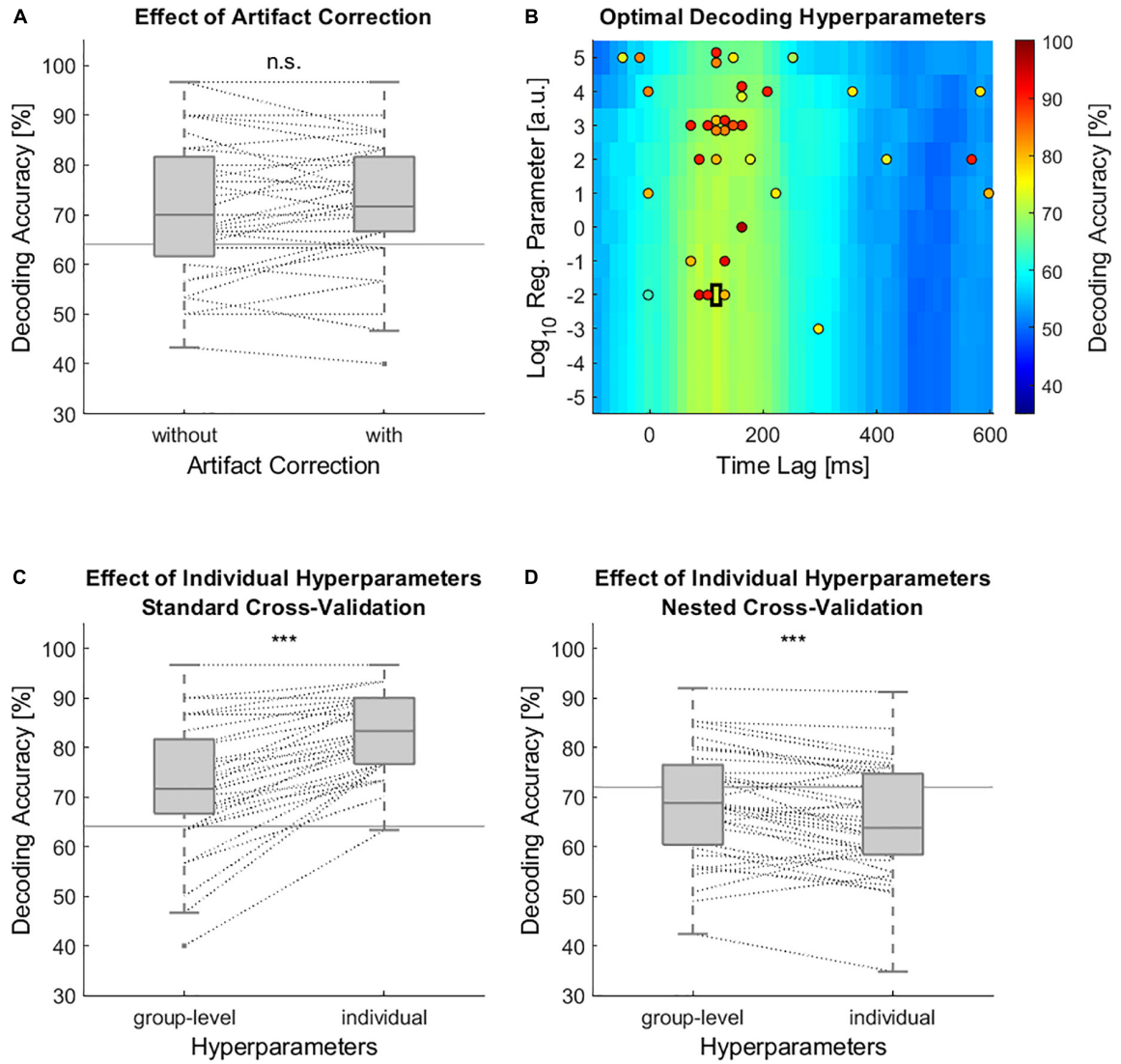


FIGURE 2 | Effects on the accuracy of speech envelope decoding models. **(A)** Decoding accuracies of the individual models with and without artifact correction. In this analysis a group-level based time lag window from 95 to 140 ms and a regularization parameter of 10^{-2} were used for all individual models. **(B)** Decoding accuracies as a function of time lag window and regularization parameter. Black rectangle marks the group-level based optimal set of hyperparameters. Colored circles mark the optimal set of hyperparameters for each participant. The color within the circle indicates the decoding accuracy of a participant which resulted from using these hyperparameters. Due to an overlap of potential time lag windows only the center of a time lag window is displayed. **(C)** Decoding accuracies with group-level chosen and individually chosen hyperparameters. These decoding accuracies were based on standard leave-one-out cross-validation including 30 test trials. **(D)** Decoding accuracy with group-level and individually chosen hyperparameters based on nested cross-validation. Within the nested cross-validation only 10 test trials were used. **(A,C,D)** Horizontal gray lines indicate chance level decoding accuracy which were based on binomial significance thresholds. Dashed lines connect data points of the same participant (n.s. non-significant, *** $p < 0.001$).

enabled us to classify to which audio book a participant was attending to **Figure 3B**. Classifying whether participants attended to the left story, using their ISC sum score with participants who attended to the left story, yielded an accuracy of 97.83%. Classifying whether participants attended to the right story, using their ISC sum score with participants who attended to the right story, yielded an accuracy of 80.05%. Both classification accuracies were clearly above chance level, which was at 65.94%. The spatial patterns of the condition-independent ISC components are shown in **Figure 3D**. We also provide the spatial patterns for all those participants attending to the left and those attending to the right in the **Supplementary Material**. Keeping in mind sign ambiguities, the spatial patterns of the left and right condition did not differ strongly from each other, nor from the condition-independent patterns (**Supplementary Figure 2**).

Spectral Entropy

Figure 4A shows the average of all individual spectrograms from 8 to 32 Hz which in turn were averaged over all but

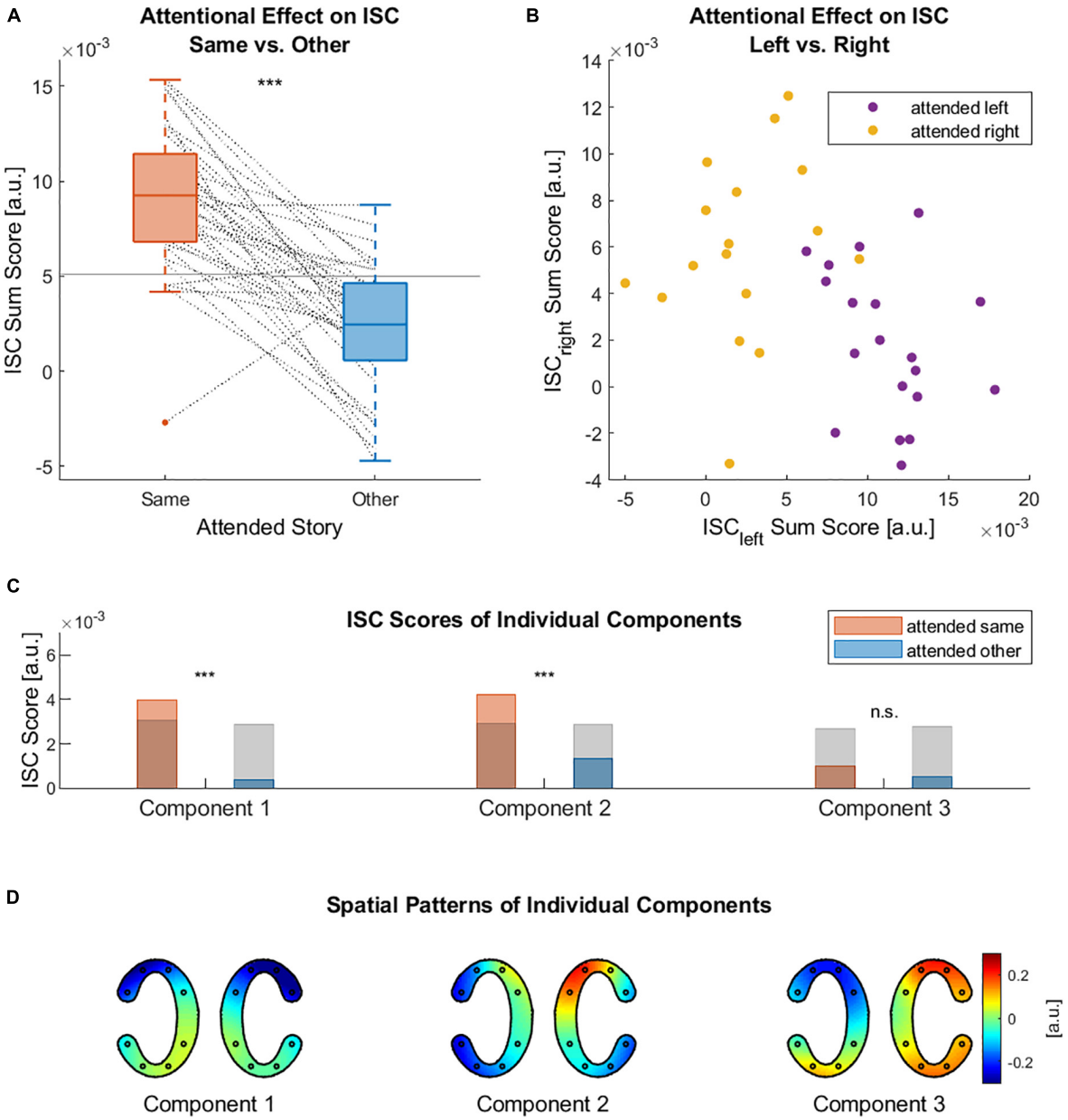


FIGURE 3 | Attentional effects on ISC. **(A)** ISC sum scores of each participant with all those attending to the same story (ISC_{same}) and with all those attending to the other story (ISC_{other}). Horizontal lines indicate chance level based on circular time-shifted data. Dashed lines connect data points of the same participant. **(B)** ISC sums scores of each participant with all those attending to the left story (ISC_{left}) and with all those attending to the right story (ISC_{right}). **(C)** Grand average of the ISC scores of three strongest components. Once computed between those participants attending to the same story and once for those attending to different stories. Gray bar indicates chance level based on circular time-shifted data. **(D)** Spatial patterns (cEEGrid topographies) of the three strongest ISC components over all participants, independent of which story they attended to. In each pair of cEEGrids the left and right cEEGrid are depicted (n.s. non-significant, *** $p < 0.001$).

the artifactual channels. The grand average spectral entropy decreased over time (**Figure 4B**, Spearman rank correlation, rho = –0.81, $p < 0.001$). On an individual level the spectral entropy significantly decreased over time for 12 participants while it significantly increased for 5 participants. For the remaining 19 participants there was no significant change over time (**Supplementary Figure 3**). In line with a decrease in the grand average spectral entropy, the grand average alpha power significantly increased over time (**Figure 4C**, Spearman rank correlation, rho = 0.82, $p < 0.001$).

Relation Between Attentional Measures

The attentional gain observed in speech envelope tracking ($Corr_{att}$–$Corr_{ign}$) correlated positively with the attentional effect observed in the ISC sum scores (ISC_{same}–ISC_{other}, **Figure 5**, Spearman rank correlation, rho = 0.3, $p = 0.04$).

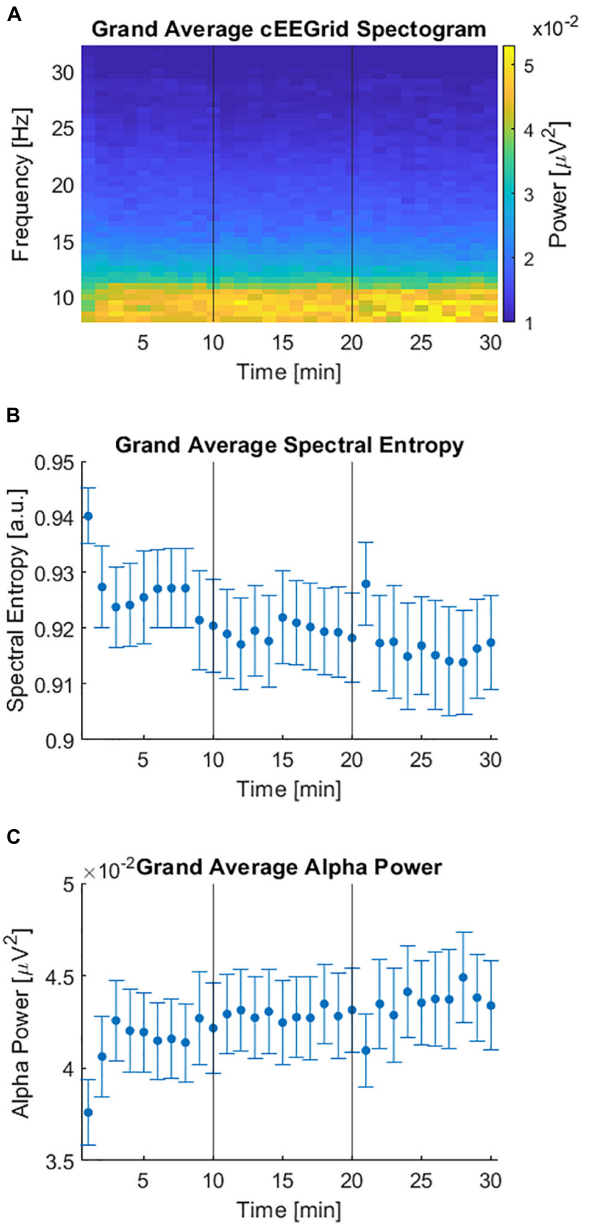


FIGURE 4 | Spectral domain of cEEGrid data during the competing speaker paradigm. **(A)** Grand average spectrogram over all channels and participants in the frequency range from 8 to 32 Hz. **(B)** Spectral entropy over time averaged over channels and participants. Error bars reflect the standard error over participants. **(C)** Alpha power (8–12 Hz) over time averaged over channels and participants. Error bars depict the standard error over participants. **(A–C)** Vertical lines at 10 and 20 min indicate the end of a preceding 10-min block.

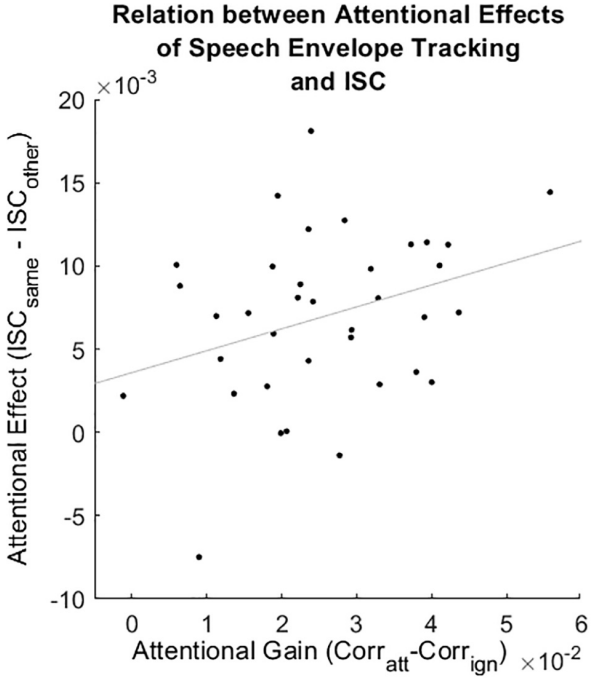


FIGURE 5 | Correlation between the attentional gain in speech envelope tracking and the attentional effect in ISC sum scores (Spearman rank correlation, rho = 0.3, $p = 0.04$). $Corr_{att}$: Spearman correlation between the predicted and the attended speech envelope. $Corr_{ign}$: Spearman correlation between the predicted and the ignored speech envelope. ISC_{same}: ISC sum score between a participant and all others attending to the same story. ISC_{other}: ISC sum score between a participant and all others attending to the other story. Gray line represents the least square regression.

There was no evidence for a relation between the attentional gain observed in speech envelope tracking and the spectral entropy, neither for any individual participant nor for the time aggregated analysis (**Supplementary Figure 4A**, Spearman rank correlation, rho = –0.22, $p = 0.19$). There was no evidence for a relation between the attentional effect observed in ISC sum scores and the spectral entropy values averaged over time (**Supplementary Figure 4B**, Spearman rank correlation, rho = 0.06, $p = 0.73$).

DISCUSSION

Methods such as speech envelope tracking, ISCs, and spectral entropy help to analyze the neural processing of continuous stimuli. We show that all three methods capture complementary information about attention when the neural data is acquired exclusively with small flex-printed electrodes placed around the ear. Speech envelope tracking reliably decodes the attended of two concurrently presented speakers using cEEGrid data. We found that artifact correction did not increase the decoding accuracies while individualizing hyperparameters of the decoding models did. Moreover, ISCs based on cEEGrid data showed more similar brain activity between an individual with those attending to the same speaker than with those attending to another speaker. Regarding spectral entropy, we found that values obtained from cEEGrid data decreased over time, potentially reflecting a decrease in the participants' level of attention. Interestingly, the attentional gain of speech envelope tracking and the attentional effect of ISC sum scores correlated positively while there was no

evidence that either of these two measures correlated with the spectral entropy values.

Speech Envelope Tracking

Effect of Artifact Correction

By approximating the analysis pipeline described in Mirkovic et al. (2016), the resulting decoding accuracies observed in the current study were comparable to Mirkovic et al. (2016). However, in contrast to our expectations, attenuating artifacts with ASR did not increase the decoding accuracy. One explanation could be that the short duration of typical artifacts only makes up a small portion of the 60 s segments that were used for decoding and thus artifact reduction does not strongly affect the decoding accuracy. In fact, only a small portion of the data contained strong artifacts which were corrected by ASR. Consequently, the decoding accuracy may benefit more from artifact correction when shorter segments of data are used. In Jaeger et al. (2020), ASR improved classification of shorter data segments, whereas Straetmans et al. (2022) showed that even for data segments as short as 5 s, decoding accuracies were not increased when the data was cleaned with ASR. We speculate that these heterogenous results could reflect the quality of the calibration data that were used to perform ASR. In Jaeger et al. (2020), the calibration data were extracted while participants performed a task (i.e., the competing speaker paradigm). In Straetmans et al. (2022), the calibration data were acquired while participants were seated without performing any task. It is known that good calibration data are crucial when performing ASR (Blum et al., 2019).

Effect of Individually Chosen Hyperparameters

We tested the effect of individualizing the classification model hyperparameters on the decoding accuracy using the commonly applied standard leave-one-out cross-validation. As could be expected, we observed higher decoding accuracies for models using individually chosen hyperparameters compared to models using group-level chosen hyperparameters. However, implementing standard leave-one-out cross-validation involves the risk of overfitting, since the same data are used for choosing the optimal hyperparameters and validating the model (Holdgraf et al., 2017). To account for this bias, we repeated the analysis using nested cross validation (Varma and Simon, 2006), where the validation of the model is done on a different part of the data than the training or the selection of hyperparameters (Parvandeh et al., 2020). In contrast to the results obtained with standard cross-validation, when implementing nested cross-validation, the models using group-level chosen hyperparameters outperformed those models using individually chosen hyperparameters. These results contradict studies showing that models fitted to the individual generally perform better than group-level based models (Mirkovic et al., 2015; O'Sullivan et al., 2015). Yet, when only a small amount of individual data is available, group-level based models outperform individualized models (Mirkovic et al., 2015). Only when a sufficient amount of data from an individual is supplied, does the individualized model outperform the group-level based model (Mirkovic et al., 2015). Therefore, we assume that a sufficiently large amount of individual data is necessary for the beneficial effect of individually chosen hyperparameters to become apparent when using nested cross-validation. Recently, a new approach has been proposed where the decoding models are initially provided with a participant-independent decoder which is then continuously updated as more data from the individual is available (Geirnaert et al., 2021a). To further investigate the effect of fitting the model to the individual, long-term recordings of an individual should be acquired. In contrast to cap-EEG acquisition, long-term data collection is certainly feasible with cEEGrids, providing good signal quality for many hours (Debener et al., 2015; Bleichner and Debener, 2017; Da Silva Souto et al., 2021; Hölle et al., 2021).

Intersubject Correlation

We provide evidence that attentional effects of EEG-based ISCs can reliably be observed even when the neural data is recorded with a small number of electrodes placed around the ear. This is not a fully independent replication of the results reported by Rosenkranz et al. (2021), as the cap-EEG analyzed in that study was simultaneously acquired with the cEEGrid data presented here. However, it shows the potential of ear-EEG to measure attentional effects of ISCs. When comparing the ISC sum scores of cap-EEG with those of cEEGrid data, it becomes apparent that the cEEGrid based ISC sum scores are less often above chance. This is also the case for the ISC scores of the individual components. The fact that fewer ISC scores are above chance for cEEGrid data may be due to the lower number of channels and their spatial coverage. In fact, cEEGrid electrodes do not cover central parts of the scalp where ISCs are most prominently expressed (Rosenkranz et al., 2021). Nevertheless, the ISC sum scores were higher in the same than in the other condition for 33 out of 36 participants. In addition, the ISC_{left} and ISC_{right} sum scores enabled us to accurately classify to which audio book a participant attended to. These results demonstrate for the first time the sensitivity of around-the-ear EEG to attentional effects in ISCs.

It has been shown that attentional effects of ISCs can also be obtained based on other physiological data such as electrodermal or heartbeat activity, yet less reliably than based on EEG (Brouwer et al., 2019; Stuldreher et al., 2020; Pérez et al., 2021). However, in terms of application, electrodermal and heartbeat activity were preferred over traditional cap-EEG as those measures are easy to apply and cost efficient. Here we show that the cEEGrid presents a suitable candidate which fulfills both criteria–it can be used to obtain reliable attentional effects in ISCs, and it can be easily applied to unobtrusively measure one's EEG. Thus, especially the combination of cEEGrids with cost-efficient data acquisition platforms such as the OpenBCI provide a setup that could be used for research in everyday life scenarios (Knierim et al., 2021). In addition, ISCs could also be based on a combination of EEG, electrodermal, and heartbeat activity, which has been shown to produce more accurate results than using EEG alone (Stuldreher et al., 2022).

Spectral Entropy

The capacity to sustain attention in demanding tasks typically declines over time, coinciding with an increase in mental fatigue

(Moore et al., 2017). Spectral entropy has been proposed as an objective measure of sustained attention (Lesenfants et al., 2018). In line with this, we found a decrease in spectral entropy over time. Since application of spectral entropy as a marker of sustained attention is a fairly new approach, there is limited evidence available to which we can compare our results. However, spectral entropy computed in the frequency range from 8 to 32 Hz strongly depends on alpha power (8–12 Hz), which in turn has also been associated with attention (Foxe and Snyder, 2011; Klimesch, 2012). The influence of alpha power on spectral entropy is evident in Lesenfants et al. (2018). They observed decreased alpha power and increased spectral entropy when participants were actively attending to a flickering stimulus compared to when the participants did not attend to the presented stimulus. We also found this inverse relation between alpha power and spectral entropy in the increase of alpha power over time. Such an increase in alpha power over time has been attributed to the depletion of attentional resources (Wascher et al., 2014). Furthermore, alpha band activity has been related to the suppression of task-irrelevant stimuli (Foxe and Snyder, 2011; Klimesch, 2012). Thus, the increase in alpha power and the decrease in spectral entropy might reflect the growing need to suppress the ignored speaker when mental fatigue accumulates.

Relation Between Attentional Measures

We found a positive relation between the attentional gain in speech envelope tracking and the attentional effect of ISCs. This suggests that both measures reflect similar phenomena. Attended stimuli evoke a stronger neural response than ignored stimuli (Picton and Hillyard, 1974). While speech envelope tracking focuses on the aspect that the neural response toward the speech envelopes is consistent over time within a participant (Aiken and Picton, 2008), ISCs focus on the aspect that the neural response toward the same external stimuli is similar between participants, at least in sensory areas (Hasson et al., 2004).

Spectral entropy on the other hand is not directly linked to the neural response to the stimulus. When computing spectral entropy in the frequency range from 8 to 32 Hz, the lower frequencies, which are relevant for the brain to track the speech envelope (Giraud and Poeppel, 2012) are neglected. Spectral entropy may rather reflect a participant's level of attention or vigilance (Lesenfants et al., 2018), that is, the capability to be aware and focus on external stimuli (van Schie et al., 2021). In contrast, speech envelope tracking and ISCs capture selective attention – the ability to select relevant and neglect irrelevant information. This may explain why we did not find a correlation between spectral entropy values and any of the two selective attention measures. This does not mean that one's level of attention or vigilance does not influence one's ability in selective attention, but only states there may not be a direct linear relation. In fact, Lesenfants and Francart (2020) showed that there is a difference in one's selective attention ability during periods of high and low levels of attention/vigilance, but the exact nature of a potential relation between the selective attention and attention/vigilance needs to be further explored.

CONCLUSION

The current study provides clear evidence that attentional measures to natural and continuous stimuli can be captured with around-the-ear EEG recordings, as provided with the cEEGrid. Ear-EEG opens up the possibility to capture neural traces of attentional processes unobtrusively in realistic everyday life scenarios. Future assistive devices could help those that have difficulties attending to one stream of information in the presence of distractor sounds.

DATA AVAILABILITY STATEMENT

The original cEEGrid contributions presented in the study are publicly available. These data can be found here: https://openneuro.org/datasets/ds004015. MATLAB code used to compute the results presented in the current study can be found on GitHub (https://doi.org/10.5281/zenodo.6379903).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Kommission für Forschungsfolgenabschätzung und Ethik, University of Oldenburg, Oldenburg, Germany. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MJ and BH performed the data acquisition. BH analyzed the data and wrote the manuscript to which MR, MJ, SD, and BM contributed with critical revisions. All authors approved the final version and agreed to be accountable for this work.

FUNDING

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy (EXC 2177/1, Project ID 390895286), and the German Federal Ministry of Education and Research (BMBF, 16SV7784). BM was funded by the DFG (Project ID 432063183).

ACKNOWLEDGMENTS

We would like to thank Lisa Straetmans for exchanging ideas regarding the effect of artifact correction on the performance of speech envelope tracking. We would further like to thank Joanna Scanlon for proof-reading the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.869426/full#supplementary-material

REFERENCES

Aiken, S. J., and Picton, T. W. (2008). Human cortical responses to the speech envelope. *Ear Hear.* 29, 139–157. doi: 10.1097/aud.0b013e31816453dc

Bleichner, M. G., and Debener, S. (2017). Concealed, Unobtrusive Ear-Centered EEG Acquisition: cEEGrids for Transparent EEG. *Front. Hum. Neurosci.* 11:163. doi: 10.3389/fnhum.2017.00163

Bleichner, M. G., Mirkovic, B., and Debener, S. (2016). Identifying auditory attention with ear-EEG: cEEGrid versus high-density cap-EEG comparison. *J. Neural. Eng* 13:66004. doi: 10.1088/1741-2560/13/6/066004

Blum, S., Mirkovic, B., and Debener, S. (2019). "Evaluation of Riemannian ASR on cEEGrid data: An artifact correction method for BCIs," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)* (Piscataway: IEEE), 3625–3630. doi: 10.1109/SMC43495.2019

Brouwer, A.-M., Stuldreher, I. V., and Thammasan, N. (2019). "Shared attention reflected in EEG, electrodermal activity and heart rate," in *Proceedings of the Workshop Socio-Affective Technologies: an interdisciplinary approach co-located with IEEE SMC 2019 (Systems, Man and Cybernetics)* (Piscataway: IEEE), 27–31. doi: 10.1088/1741-2552/aba87d

Chang, C. Y., Hsu, S. H., Pion-Tonachini, L., and Jung, T. P. (2018). "Evaluation of artifact subspace reconstruction for automatic EEG artifact removal," in *In 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Piscataway: IEEE), 1242–1245.

Cohen, S. S., and Parra, L. C. (2016). Memorable Audiovisual Narratives Synchronize Sensory and Supramodal Neural Responses. *eNeuro* 3, ENEURO.0203-16.2016. doi: 10.1523/ENEURO.0203-16.2016

Crosse, M. J., Di Liberto, G. M., Bednar, A., and Lalor, E. C. (2016). The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. *Front. Hum. Neurosci.* 10:604. doi: 10.3389/fnhum.2016.00604

Da Silva Souto, C. F., Pätzold, W., Wolf, K. I., Paul, M., Matthiesen, I., et al. (2021). Flex-Printed Ear-EEG Sensors for Adequate Sleep Staging at Home. *Front. Digit. Health* 3:688122. doi: 10.3389/fdgth.2021.688122

Debener, S., Emkes, R., Vos, M., de, and Bleichner, M. (2015). Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear. *Sci. Rep.* 5:16743. doi: 10.1038/srep16743

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

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

Dmochowski, J. P., Sajda, P., Dias, J., and Parra, L. C. (2012). Correlated components of ongoing EEG point to emotionally laden attention - a possible marker of engagement? *Front. Hum. Neurosci.* 6:112. doi: 10.3389/fnhum.2012.00112

Foxe, J. J., and Snyder, A. C. (2011). The Role of Alpha-Band Brain Oscillations as a Sensory Suppression Mechanism during Selective Attention. *Front. Psychol.* 2:154. doi: 10.3389/fpsyg.2011.00154

Geirnaert, S., Francart, T., and Bertrand, A. (2021a). Unsupervised Self-Adaptive Auditory Attention Decoding. *IEEE J. Biomed. Health Inform.* 25, 3955–3966. doi: 10.1109/JBHI.2021.3075631

Geirnaert, S., Vandecappelle, S., Alickovic, E., Cheveigne, A., de, Lalor, E., et al. (2021b). Electroencephalography-Based Auditory Attention Decoding: Toward Neurosteered Hearing Devices. *IEEE Signal Process. Mag.* 38, 89–102. doi: 10.1109/MSP.2021.3075932

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

Hamilton, L. S., and Huth, A. G. (2020). The revolution will not be controlled: natural stimuli in speech neuroscience. *Lang. Cogn. Neurosci.* 35, 573–582. doi: 10.1080/23273798.2018.1499946

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject Synchronization of Cortical Activity During Natural Vision. *Science* 303, 1634–1640. doi: 10.1126/science.1089506

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., et al. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage* 87, 96–110. doi: 10.1016/j.neuroimage.2013.10.067

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

Hölle, D., Meekes, J., and Bleichner, M. G. (2021). Mobile ear-EEG to study auditory attention in everyday life : Auditory attention in everyday life. *Behav. Res. Methods* 53, 2025–2036. doi: 10.3758/s13428-021-01538-0

Holmes, E., and Griffiths, T. D. (2019). 'Normal' hearing thresholds and fundamental auditory grouping processes predict difficulties with speech-in-noise perception. *Sci Rep.* 9, 16771. doi: 10.1038/s41598-019-53353-5

Holtze, B., Jaeger, M., Debener, S., Adiloğlu, K., and Mirkovic, B. (2021). Are They Calling My Name? Attention Capture Is Reflected in the Neural Tracking of Attended and Ignored Speech. *Front. Neurosci.* 15:643705. doi: 10.3389/fnins.2021.643705

Jaeger, M., Mirkovic, B., Bleichner, M. G., and Debener, S. (2020). Decoding the Attended Speaker From EEG Using Adaptive Evaluation Intervals Captures Fluctuations in Attentional Listening. *Front. Neurosci.* 14:603. doi: 10.3389/fnins.2020.00603

Janssen, T. W., Grammer, J. K., Bleichner, M. G., Bulgarelli, C., Davidesco, I., Dikker, S., et al. (2021). Opportunities and Limitations of Mobile Neuroimaging Technologies in Educational Neuroscience. *Mind Brain Educ.* 15, 354–370. doi: 10.1111/mbe.12302

Kang, S. S., Lano, T. J., and Sponheim, S. R. (2015). Distortions in EEG interregional phase synchrony by spherical spline interpolation: causes and remedies. *Neuropsychiatr. Electrophysiol.* 1:9. doi: 10.1186/s40810-015-0009-5

Kayser, H., Ewert, S. D., Anemüller, J., Rohdenburg, T., Hohmann, V., and Kollmeier, B. (2009). Database of Multichannel In-Ear and Behind-the-Ear Head-Related and Binaural Room Impulse Responses. *EURASIP J. Adv. Signal Process.* 2009:157. doi: 10.1155/2009/298605

Ki, J. J., Kelly, S. P., and Parra, L. C. (2016). Attention Strongly Modulates Reliability of Neural Responses to Naturalistic Narrative Stimuli. *J. Neurosci.* 36, 3092–3101. doi: 10.1523/JNEUROSCI.2942-15.2016

Klimesch, W. (2012). α-band oscillations, attention, and controlled access to stored information. *Trends Cogn. Sci.* 16, 606–617. doi: 10.1016/j.tics.2012.10.007

Knierim, M. T., Berger, C., and Reali, P. (2021). Open-source concealed EEG data collection for Brain-computer-interfaces - neural observation through OpenBCI amplifiers with around-the-ear cEEGrid electrodes. *Brain Comput. Interfaces* 8, 161–179. doi: 10.1080/2326263X.2021.1972633

Lauter, J. L., and Karzon, R. G. (1990). Individual differences in auditory electric responses: comparisons of between-subject and within-subject variability. IV. Latency-variability comparisons in early, middle, and late responses. *Scand. Audiol.* 19, 175–182. doi: 10.3109/01050399009070769

Lesenfants, D., and Francart, T. (2020). The interplay of top-down focal attention and the cortical tracking of speech. *Sci. Rep.* 10:6922. doi: 10.1038/s41598-020-63587-3

Lesenfants, D., Habbal, D., Chatelle, C., Soddu, A., Laureys, S., and Noirhomme, Q. (2018). Toward an Attention-Based Diagnostic Tool for Patients With Locked-in Syndrome. *Clin. EEG Neurosci.* 49, 122–135. doi: 10.1177/1550059416674842

Looney, D., Kidmose, P., Park, C., Ungstrup, M., Rank, M., Rosenkranz, K., et al. (2012). The in-the-ear recording concept: user-centered and wearable brain monitoring. *IEEE Pulse* 3, 32–42. doi: 10.1109/MPUL.2012.2216717

Meiser, A., Tadel, F., Debener, S., and Bleichner, M. G. (2020). The Sensitivity of Ear-EEG: Evaluating the Source-Sensor Relationship Using Forward Modeling. *Brain Topogr.* 33, 665–676. doi: 10.1007/s10548-020-00793-2

Mirkovic, B., Bleichner, M. G., Vos, M., de, and Debener, S. (2016). Target Speaker Detection with Concealed EEG Around the Ear. *Front. Neurosci.* 10:349. doi: 10.3389/fnins.2016.00349

Mirkovic, B., Debener, S., Jaeger, M., Vos, M., and de. (2015). Decoding the attended speech stream with multi-channel EEG: implications for online, daily-life applications. *J. Neural. Eng.* 12:46007. doi: 10.1088/1741-2560/12/4/046007

Mirkovic, B., Debener, S., Schmidt, J., Jaeger, M., and Neher, T. (2019). Effects of directional sound processing and listener's motivation on EEG responses to continuous noisy speech: Do normal-hearing and aided hearing-impaired listeners differ? *Hear Res.* 377, 260–270. doi: 10.1016/j.heares.2019.04.005

Moore, T. M., Key, A. P., Thelen, A., and Hornsby, B. W. Y. (2017). Neural mechanisms of mental fatigue elicited by sustained auditory processing. *Neuropsychologia* 106, 371–382. doi: 10.1016/j.neuropsychologia.2017.10.025

Mullen, T. R., Kothe, C. A. E., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., et al. (2015). Real-Time Neuroimaging and Cognitive Monitoring Using Wearable Dry EEG. *IEEE Trans. Biomed. Eng.* 62, 2553–2567. doi: 10.1109/TBME.2015.2481482

Nastase, S. A., Gazzola, V., Hasson, U., and Keysers, C. (2019). Measuring shared responses across subjects using intersubject correlation. *Soc. Cogn. Affect Neurosci.* 14, 667–685. doi: 10.1093/scan/nsz037

Nogueira, W., Dolhopiatenko, H., Schierholz, I., Büchner, A., Mirkovic, B., Bleichner, M. G., et al. (2019). Decoding Selective Attention in Normal Hearing Listeners and Bilateral Cochlear Implant Users With Concealed Ear EEG. *Front. Neurosci.* 13:720. doi: 10.3389/fnins.2019.00720

O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., et al. (2015). Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cereb. Cortex* 25, 1697–1706. doi: 10.1093/cercor/bht355

Parra, L. C., Haufe, S., and Dmochowski, J. P. (2018). Correlated components analysis-extracting reliable dimensions in multivariate data. [Preprint] *Bioarxiv* doi: 10.48550/arXiv.1801.08881

Parvandeh, S., Yeh, H.-W., Paulus, M. P., and McKinney, B. A. (2020). Consensus features nested cross-validation. *Bioinformatics* 36, 3093–3098. doi: 10.1093/bioinformatics/btaa046

Pérez, P., Madsen, J., Banellis, L., Türker, B., Raimondo, F., Perlbarg, V., et al. (2021). Conscious processing of narrative stimuli synchronizes heart rate between individuals. *Cell Rep.* 36:109692. doi: 10.1016/j.celrep.2021.109692

Picton, T. W., and Hillyard, S. A. (1974). Human auditory evoked potentials. II. Effects of attention. *Electroencephalogr. Clin. Neurophysiol.* 191–199. doi: 10.1016/0013-4694(74)90156-4

Poulsen, A. T., Kamronn, S., Dmochowski, J., Parra, L. C., and Hansen, L. K. (2017). EEG in the classroom: Synchronised neural recordings during video presentation. *Sci. Rep.* 7:43916. doi: 10.1038/srep43916

Power, A. J., Foxe, J. J., Forde, E.-J., Reilly, R. B., and Lalor, E. C. (2012). At what time is the cocktail party? A late locus of selective attention to natural speech. *Eur. J. Neurosci.* 35, 1497–1503. doi: 10.1111/j.1460-9568.2012.08060.x

Rosenkranz, M., Holtze, B., Jaeger, M., and Debener, S. (2021). EEG-Based Intersubject Correlations Reflect Selective Attention in a Competing Speaker Scenario. *Front. Neurosci.* 15:685774. doi: 10.3389/fnins.2021.685774

Shinn-Cunningham, B. G., and Best, V. (2008). Selective attention in normal and impaired hearing. *Trends Amplif.* 12, 283–299. doi: 10.1177/1084713808325306

Stone, M. (1974). Cross-Validation and Multinomial Prediction. *Biometrika* 61, 509–515. doi: 10.1093/biomet/61.3.509

Straetmans, L., Holtze, B., Debener, S., Jaeger, M., and Mirkovic, B. (2022). Neural tracking to go: auditory attention decoding and saliency detection with mobile EEG. *J. Neural. Eng.* 18:685774. doi: 10.1088/1741-2552/ac42b5

Stuldreher, I. V., Merasli, A., Thammasan, N., van Erp, J. B. F., and Brouwer, A.-M. (2022). Unsupervised Clustering of Individuals Sharing Selective Attentional Focus Using Physiological Synchrony. *Front. Neuroergonomics* 2:750248. doi: 10.3389/fnrgo.2021.750248

Stuldreher, I. V., Thammasan, N., van Erp, J. B. F., and Brouwer, A.-M. (2020). Physiological synchrony in EEG, electrodermal activity and heart rate reflects shared selective auditory attention. *J. Neural. Eng.* 17:46028.

van Schie, M. K. M., Lammers, G. J., Fronczek, R., Middelkoop, H. A. M., and van Dijk, J. G. (2021). Vigilance: discussion of related concepts and proposal for a definition. *Sleep Med.* 83, 175–181. doi: 10.1016/j.sleep.2021.04.038

Varma, S., and Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* 7:91. doi: 10.1186/1471-2105-7-91

Viertiö-Oja, H., Maja, V., Särkelä, M., Talja, P., Tenkanen, N., Tolvanen-Laakso, H., et al. (2004). Description of the Entropy algorithm as applied in the Datex-Ohmeda S/5 Entropy Module. *Acta Anaesthesiol. Scand.* 48, 154–161. doi: 10.1111/j.0001-5172.2004.00322.x

Wascher, E., Rasch, B., Sänger, J., Hoffmann, S., Schneider, D., Rinkenauer, G., et al. (2014). Frontal theta activity reflects distinct aspects of mental fatigue. *Biol. Psychol.* 96, 57–65. doi: 10.1016/j.biopsycho.2013.11.010

Wong, D. D. E., Fuglsang, S. A., Hjortkjær, J., Ceolini, E., Slaney, M., and Cheveigné, A. D. (2018). A Comparison of Regularization Methods in Forward and Backward Models for Auditory Attention Decoding. *Front. Neurosci.* 12:531. doi: 10.3389/fnins.2018.00531

Zhang, D. (2018). "Computational EEG Analysis for Hyperscanning and Social Neuroscience," in *Computational EEG Analysis*, ed. C.-H. Im (Singapore: Springer Singapore), 215–228. doi: 10.1007/978-981-13-0908-3_10

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

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

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 | Frontiers in Neuroscience

ORIGINAL RESEARCH
published: 24 June 2022
doi: 10.3389/fnins.2022.883966

Pre-gelled Electrode Grid for Self-Applied EEG Sleep Monitoring at Home

Carlos F. da Silva Souto[1], Wiebke Pätzold[1], Marina Paul[2], Stefan Debener[1,3,4] and Karen Insa Wolf[1]*

[1] Fraunhofer Institute for Digital Media Technology IDMT, Branch Hearing, Speech and Audio Technology HSA, Oldenburg, Germany, [2] PSG-Auswertungs-Service, Stadtlohn, Germany, [3] Neuropsychology Laboratory, Department of Psychology, University of Oldenburg, Oldenburg, Germany, [4] Cluster of Excellence Hearing4All, University of Oldenburg, Oldenburg, Germany

OPEN ACCESS

Edited by:
Jérémie Voix, École de Technologie Supérieure (ÉTS), Canada

Reviewed by:
Arun Sasidharan, National Institute of Mental Health and Neurosciences, India
Bertille Somon, Université de Toulouse, France

***Correspondence:**
Carlos F. da Silva Souto
carlos.filipe.da.silva.souto@idmt.fraunhofer.de

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *25 February 2022*
Accepted: *06 June 2022*
Published: *24 June 2022*

Citation:
da Silva Souto CF, Pätzold W, Paul M, Debener S and Wolf KI (2022) Pre-gelled Electrode Grid for Self-Applied EEG Sleep Monitoring at Home. Front. Neurosci. 16:883966. doi: 10.3389/fnins.2022.883966

The need for diagnostic capabilities for sleep disorders such as sleep apnea and insomnia far exceeds the capacity of inpatient sleep laboratories. Some home monitoring systems omit electroencephalography (EEG) because trained personnel may be needed to apply EEG sensors. Since EEG is essential for the detailed evaluation of sleep, better systems supporting the convenient and robust recording of sleep EEG at home are desirable. Recent advances in EEG acquisition with flex-printed sensors promise easier application of EEG sensor arrays for chronic recordings, yet these sensor arrays were not designed for sleep EEG. Here we explored the self-applicability of a new sleep EEG sensor array (trEEGrid) without prior training. We developed a prototype with pre-gelled neonatal ECG electrodes placed on a self-adhesive grid shape that guided the fast and correct positioning of a total of nine electrodes on the face and around the ear. Positioning of the sensors was based on the results of a previous ear-EEG sleep study (da Silva Souto et al., 2021), and included electrodes around the ear, one eye, and the chin. For comparison, EEG and electrooculogram channels placed according to the American Academy of Sleep Medicine criteria, as well as respiratory inductance plethysmography on thorax and abdomen, oxygen saturation, pulse and body position were included with a mobile polysomnography (PSG) system. Two studies with 32 individuals were conducted to compare the signal quality of the proposed flex-printed grid with PSG signals and to explore self-application of the new grid at home. Results indicate that the new array is self-applicable by healthy participants without on-site hands-on support. A comparison of the hypnogram annotations obtained from the data of both systems revealed an overall substantial agreement on a group level (Cohen's $\kappa = 0.70 \pm 0.01$). These results suggest that flex-printed pre-gelled sensor arrays designed for sleep EEG acquisition can facilitate self-recording at home.

Keywords: EEG, sleep staging, grapho-elements, flexible PCB electrodes, self-application

INTRODUCTION

Sleep disorders are on the rise (Jahrami et al., 2021). Traditionally, a sleep laboratory has been the best diagnostic tool to identify symptoms concerning sleep. However, many sleep disorders are evasive to one night of disrupted sleep in a laboratory setting, which may lead to over- or underdiagnosing of symptoms. Typical sleep disorders like insomnia or sleep bruxism may not be correctly identified in a single night of sleep, but only over time (Miettinen et al., 2017). Similarly, gradual changes in sleep architecture associated with neurological diseases like Parkinson's disease (Figorilli et al., 2020) or Alzheimer's disease (Mander, 2020), or mental disorders like depression (Paunio et al., 2015; Goldstone et al., 2020) cannot be detected easily in an environment that is set up for a finite diagnosis after one night. Reliable diagnostic tools are needed that can be used long-term at home with minimal invasion to a patient's comfort and rest.

If the sleep laboratory is to be considered the highest level of sleep monitoring (type I sleep measurement, typification based on Miettinen et al., 2017), the next level is the preparation of sensors at a sleep center, after which patients are sent home to sleep (type II sleep measurement). While this second level may give professionals more control over data quality, having participants or patients come into a center to prepare for unattended at-home measurements has shown failure rates that are comparable to self-applied unattended at-home measurements (type III sleep measurement). Reliable type III sleep measurements are desirable, however. In the light of the ongoing Covid-19 pandemic, the focus of health monitoring has been shifted even more to the need for self-applicability of sensor systems. In addition to enabling repeated measurements and long-term monitoring, systems suited for self-application add more possibilities to reach remote, bedridden, or quarantined individuals. Self-application also adds convenience for the patient and can therefore increase compliance with the procedure, especially if multiple recordings are planned (Mikkelsen et al., 2019). Ecological validity is further increased because patients can implement the preparation of the measurement into their nighttime routine and the quality of sleep is known to be better when individuals can sleep in familiar environments (e.g., Tamaki et al., 2016).

In this paper, we will focus on two aspects of self-application, data validity and signal quality. Data validity concerns the prospect of deriving information suitable for the relevant interpretation or diagnosis from the data. Signal quality concerns all aspects regarding the usefulness of the data in general, including impedance levels, the number of channels that delivered stable recordings and the signal-to-noise ratio. Sleep staging, for instance, can only be achieved if electrodes are placed at appropriate scalp positions (data validity) and the recorded data offers sufficient information for an adequate interpretation (signal quality). In a first feasibility study, we validate adequate electrodes positioning on the scalp for sleep staging. In a second study, participants were instructed to find the correct positioning of a sensor array themselves. This latter aspect of self-application concerns signal quality in particular. Previous EEG sensors designed for self-application have used dry electrodes (i.e., Arnal et al., 2019). While dry electrodes require no or little skin preparation and can be re-arranged, which makes them convenient for self-application, they require physical pressure to hold electrodes in place, causing discomfort, headaches and sleep disruption. In laboratory settings, wet electrodes remain the gold standard when the aim is to collect data with optimal signal quality. Professionals use wet electrodes in combination with different preparation techniques (skin cleansing and abrasion) and substances (conductive gels or pastes, special adhesives). In the context of self-application, individuals must be trained to perform these or similar steps themselves.

Previously, we showed that a system consisting of a single flexible electrode patch around the ear, a cEEGrid (Debener et al., 2015), delivers enough information for a satisfactory sleep analysis (da Silva Souto et al., 2021). However, results were significantly more reliable if EOG information from two electrodes near the eyes were included in the information available to the sleep analyst. Based on these results, we designed a new electrodes array to include positions behind the ear, around the eye and on the chin, for additional EMG acquisition (**Figure 1**). The new configuration omitted the area below the ear in favor of a bridge above the ear to comfortably connect electrodes from the facial area to those located behind the ear. Preliminary exploration revealed that this design reduced discomfort that can be associated with talking or eating with cEEGrids. Furthermore, this bridge facilitates self-application, as the ear provides additional temporary support. A near-ear electrode in the facial area close to the tragus substitutes the bottom end electrode of the cEEGrid. Concerning grapho-elements, our previous study showed that sleep spindles and K-complexes recorded at Fpz are best represented in cEEGrid channel combinations that point in the direction of Fpz. Here, we expanded the principle of linear combinations by configuring facial electrodes to allow for directional analysis pointing not only to frontal areas, but to occipital and central areas as well. This allowed us to focus on frontal areas that are especially relevant for sleep staging but also offer applications beyond sleep.

In the first study, we validated a miniaturized EEG setup consisting of the new electrode constellation and a wearable amplifier against a commercial mobile PSG system (*validation study*). In the second study, we tested the self-applicability of this system by giving it to participants and instructing them remotely to prepare the sensors for a measurement (*self-application study).*

MATERIALS AND METHODS

Measurement Systems

The new electrode grid introduced will be referred to as *trEEGrid* system and was used in combination with a wireless EEG amplifier (Smarting Sleep, mBrainTrain, Serbia). The *trEEGrid* is based on experience with the cEEGrid ear-EEG array (**Figure 1A**), but specifically designed to foster self-application for sleep EEG acquisition at home. In study 1, the general

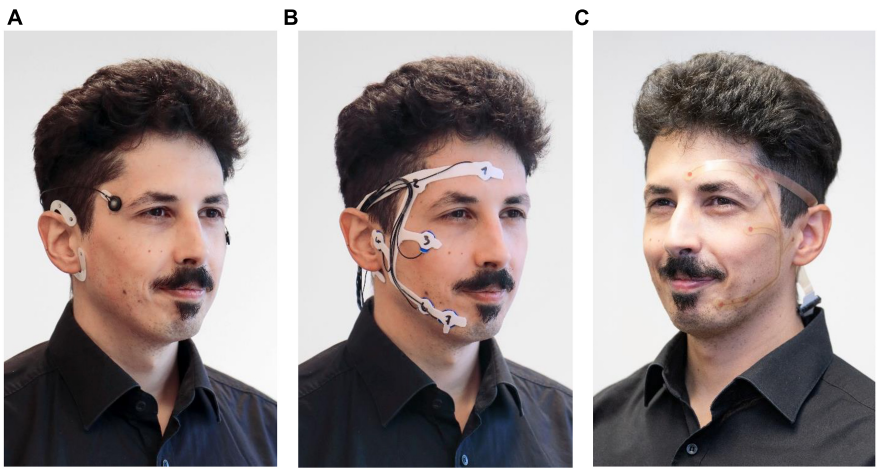


FIGURE 1 | Development steps of the self-applicable, pre-gelled *trEEGrid*. **(A)** stage 1: cEEGrid + EOG, da Silva Souto et al. (2021). **(B)** stage 2: foam *trEEGrid*, current study. **(C)** stage 3: *trEEGrid* prototype on a flexible PCB, future development. ©Fraunhofer IDMT/Anika Bödecker.

design is tested using a prototype built from pre-gelled neonatal ECG electrodes. The electrodes were embedded within a self-adhesive grid shape using a medical foam plaster (**Figure 1B**). This prototype helped to validate finalize the *trEEGrid* design, as shown on a PCB grid in **Figure 1C**. We refer to this intermediate design as foam *trEEGrid*. It included nine single-use self-adhesive gel electrodes (seven channels for EEG, EOG and EMG analysis, reference and ground).

The new prototype was validated by concurrent recordings with commercial PSG system (SOMNOscreen Plus, SomnoMedics, Germany; further referred to as *PSG system*). The commercial system included six EEG electrodes, two piezosensoric belts for thorax and abdomen expansion and a finger clip sensor, which measured oxygen saturation and pulse. The two EMG electrodes of the *PSG system* were not used, since they competed with the space of the *trEEGrid* EMG channels R6 and R7. Electrodes were gold-plated cup electrodes mounted with adhesive gel.

Participants

Twenty healthy individuals participated in the *validation study*. Eight datasets had to be excluded. Six datasets were excluded due to technical problems concerning the *trEEGrid*, one further dataset due to instability of the Bluetooth connection to the amplifier, and another dataset due to power failure of the SOMNOscreen *PSG system*. The 12 participants (three females, nine males, mean age = 28.9 years, range = 18–45 years) entering the analysis reported no sleep disorders. Each participant provided one night of sleep data, recorded during sleep at home.

Twelve further individuals participated in the *self-application study* (four females, eight males, mean age = 28.8 years, range = 18–55 years). No participant had prior experience in either handling or applying the foam *trEEGrid*, nor was involved in its design process. All participants followed instruction for self-applicating the *trEEGrid* system described above and succeeded in providing a 30 min EEG recording.

For both studies, written consent was given by all participants. Both studies were approved by the local ethics board.

Data Acquisition

Validation Study

For the *validation study*, participants came into the lab in the early evening to be prepared for the study. After given written consent, they were outfitted with the two systems, the commercial *PSG system* and the *trEEGrid* system.

The sensors of the *PSG system* (six EEG electrodes, two piezosensoric belts for thorax and abdomen expansion and a finger clip sensor for oxygen saturation and pulse measurement) were connected to a recording box worn in front of the torso on a belt. EEG recording was sampled at 256 Hz with a reference at Cz and ground at Fz.

The *trEEGrid* electrodes were connected via cable to a smarting amplifier worn in a pouch on a shoulder strap and recording at a sampling rate of 250 Hz to a raspberry pi-based recording station via Bluetooth (see **Figure 2A** for setup). Participants were then sent home to sleep. The commercial system started recording at a pre-set time indicated as just prior to bedtime by the participant. The *trEEGrid* system was manually started by the participant on a smartphone app before turning off the light. Impedance current was active during recording for continuous impedance measurements. The next morning, participants manually stopped the recording and took off both systems without assistance.

Self-Application Study

For the *self-application study*, participants either picked up the materials for the study and joined the online, supervised instruction session from home (10 participants) or they were placed alone in a room at the institute with materials and joined the online, supervised instruction session on a laptop set up for them (two participants). Materials included the *trEEGrid* system, consisting of the self-applicable electrode grid and the amplifier

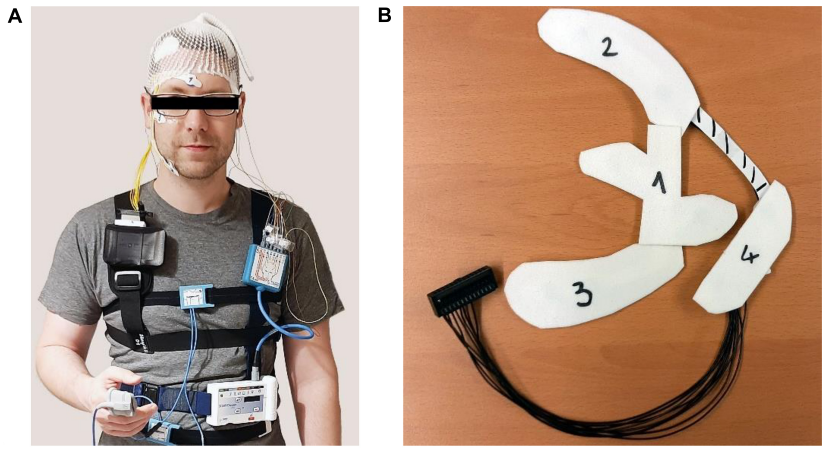


FIGURE 2 | (A) Recording setup of the *validation study* including the commercial at-home *PSG system* and the *trEEGrid* system worn at the same time. **(B)** View on foam electrode grid prepared for self-application with wax paper covers in place to guide the participant through application.

already attached to the chest belt, a mirror, and four cotton pads soaked in alcohol. The *trEEGrid* system was prepared for use by removing adhesives from both the electrodes and the foam grid structure and replacing it by four fitted pieces of wax paper numbered 1 through 4 in the order of the planned removal (see **Figure 2B**). The experimenter started the online session at the appointed time and taught the participants through a detailed picture presentation consisting of 22 slides accompanied by oral instruction how to self-apply the electrode grid, leaving room for individual questions if needed. In order, participants were instructed to put on the chest belt, prepare the skin, remove the pieces of wax paper and self-apply the electrode grid section by section and connect the grid to the amplifier. Time was recorded for this procedure. Participants filled in a questionnaire, rating statements concerning aspects of self-applicability on an agreement scale of 1 to 5 (1 = "very much," 2 = "rather," 3 = "neutral," 4 = "slightly," 5 = "not at all") and answered additional questions in a free text format upon completion of the study.

Data Analysis of Validation Study

Data Preprocessing

Data from both systems (*trEEGrid* and *PSG system*) were prepared for the annotation and the correlation analysis as followed.

First EEG signals were bandstop filtered from 45 to 55 Hz, (phase true, 4th order Butterworth) to reduce the 50 Hz line noise. EEG data of *trEEGrid* was additionally bandstop filtered from 60 to 65 Hz, (phase true, 4th order Butterworth) to reduce impedance current. EEG data of *trEEGrid* and *PSG system* were bandpass filtered from 0.5 to 40 Hz, (phase true, 4th order Butterworth) to reduce electrode drift and noise containing high-frequency components. Due to the Bluetooth connection used in the *trEEGrid* system, its data was checked for package loss using the recorded package number information. If necessary, zeros were added where packages were lost (overall less than 0.0007% of all packages were lost). Data of both systems were downsampled to 125 Hz and then synchronized by aligning ten eyeblinks at the start and end of each measurement. The excess samples before the first and after the last aligned eyeblink were discarded. If there was a size difference of both datasets due to diverging system clocks, *PSG system* data was resized to fit to *trEEGrid* data.

Channels of both datasets were selected, re-referenced and labeled as listed in **Table 1**. For the *trEEGrid* layout, linear combinations of channels were used. This step was motivated to approximately represent the EEG measured at classic PSG-relevant scalp positions Fpz, C4 and O2 when referenced to the right mastoid process M2 (in the following marked with an asterisk, as shown in **Figure 3**).

TABLE 1 | Channel configuration of the *trEEGrid* and *PSG system* data sets.

Data set	*trEEGrid*	PSG system
EEG channels	R1 ➔ *Fpz_M2	F3_M2
	R5 ➔ *C4_M2	F4_M1
	R5-R4 ➔ *O2_M2	C3_M2
		C4_M1
		O3_M2
		O4_M1
EOG	R1-R2 ➔ *EOGh	E1_M2
	R2-R3 ➔ *EOGv	E2_M1
	R1-R3 ➔ *EOGd	
EMG	R7-R6 ➔ *EMG	None
Additional sensors	None	SPO2
		Puls
		Pleth
		RIP_Thrx
		RIP_Abdom
		Summe_RIPs

*The symbol * is used to differ the names of the channel-combinations from ones of the classical scalp positions.*
This is done because the channel-combinations are used as an estimation of the EEG that could be measured at those classical scalp positions.

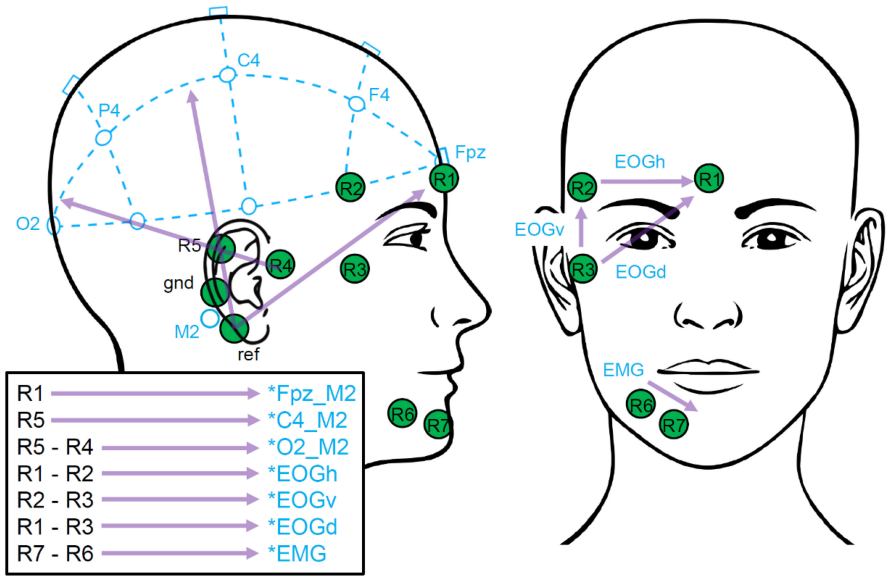


FIGURE 3 | *trEEGrid* channel combinations. These bipolar channels extrapolate toward classical PSG scalp positions (Fpz, C4, O2) as well as EOG and EMG, as illustrated.

Data Annotation of the Expert Scorer

An expert polysomnographic technologist with 15 years of experience polysomnography (MP, subsequently referred to as expert scorer) annotated the sleep data of both systems independently, by evaluating consecutive 30 s segments. Participant labels of both systems were randomized independently prior annotation to lower a possible bias by the expert scorer. An open source, sleep analysis software was used for this purpose (Combrisson et al., 2017). Sleep staging was done in accordance with AASM guidelines [annotated stages: awake (W), N1, N2, N3 and REM (rapid eye movement)]. In addition, sleep spindles were annotated in EEG data of the *PSG system*.

Statistical Analysis

To statistically evaluate differences of the impedance measurements, a repeated analysis of variance (rANOVA) was done. To test for sphericity of the data a Mauchly-Test was applied. In case of significance, results were Greenhouse-Geisser corrected.

To statistically evaluate the accordance of the different hypnograms, Cohen's Kappa was used to determine inter-rater reliability (agreement scale according to Landis and Koch, 1977). For a calculation of the reliability over participants, the corresponding hypnograms were concatenated. To this end, confusion matrices of the different hypnograms were calculated and used to determine Cohen's Kappa. Cohen's Kappa was calculated for every single sleep stage vs. the rest, and across all sleep stages.

Correlation Analysis

da Silva Souto et al. (2021) showed that sub-hairline EEG electrodes can be used to estimate the EEG measured at further distanced location of the scalp. The authors correlated sleep spindles annotated at a classical Fpz electrode with corresponding EEG recordings of linear combinations of ear electrodes and showed that highest correlations were scored at channel combinations that point in the direction of Fpz. To further explore this approach, we expanded the analysis to central and occipital PSG locations of the scalp. Therefore, we first extracted spindle events by separating of the transient EEG of the *trEEGrid* channels (R1, . . ., R7) as well as the transient EEG of the PSG channels F4_M2, C4_M2 and O2_M2 into epochs according to the sleep spindle annotations of the expert scorer in *PSG system* rating. The annotated epochs of each channel F4_M2, C4_M2 and O2_M2 are used as reference signals in three appropriate test conditions of the correlation analysis. DC-offset was reduced by subtracting the average over all samples of the respective epoch. Finally, correlation coefficients were calculated between the epochs of the three reference signals and the respective transient data of all possible single channel and bipolar channel combinations of *trEEGrid* electrodes (28 in total) as shown in the following:

*R*1,
*R*1-*R*2, *R*2,
*R*1-*R*3, *R*2-*R*3, *R*3,
*R*1-*R*4, *R*2-*R*4, *R*3-*R*4, *R*4,
*R*1-*R*5, *R*2-*R*5, *R*3-*R*5, *R*4-*R*5, *R*5,
*R*1-*R*6, *R*2-*R*6, *R*3-*R*6, *R*4-*R*6, *R*5-*R*6, *R*6,
*R*1-*R*7, *R*2-*R*7, *R*3-*R*7, *R*4-*R*7, *R*5-*R*7, *R*6-*R*7, *R*7.

If a *trEEGrid* channel impedance exceeded 60 kΩ during an annotation window, the results of this channel and all related channel combinations were discarded for this specific epoch (13.6% epochs rejected on average for all listed combination and

TABLE 2 | *trEEGrid* system impedances at the start and end of night measurements in the *validation study* in kΩ (impedances above 100 kΩ are marked as "-").

Participants	Impedance measurements [kΩ]													
	R1		R2		R3		R4		R5		R6		R7	
	Start	End	Start	End	Start	End	Start	End	Start	End	Start	End	Start	End
P1	7.2	11.7	17.9	15.5	10.9	10.0	–	45.1	16.1	10.8	–	3.9	8.1	8.3
P2	13.1	14.0	14.5	11.8	40.1	25.3	–	22.8	55.2	44.1	–	36.7	–	34.5
P3	12.4	9.7	38.4	26.3	19.6	13.3	36.6	28.5	93.2	54.0	23.0	16.5	21.6	12.3
P4	17.4	21.9	18.5	24.4	22.8	5.1	24.6	34.2	18.4	17.7	18.1	25.3	17.6	19.1
P5	18.8	13.2	29.4	11.8	28.6	24.4	51.8	19.5	39.2	19.9	14.7	12.8	22.3	–
P6	55.1	25.0	52.2	22.5	56.3	51.5	76.8	54.0	62.5	31.9	–	–	–	–
P7	13.1	10.9	16.5	13.0	25.8	17.6	–	18.2	30.4	21.3	–	–	–	–
P8	11.6	12.2	16.2	12.5	15.1	14.4	18.9	12.0	4.9	5.0	27.2	41.9	20.4	12.5
P9	24.1	37.9	33.2	33.2	21.7	13.1	–	36.3	53.5	40.3	–	48.2	52.1	28.5
P10	11.0	15.1	15.0	–	15.0	14.3	27.7	23.9	–	31.5	15.8	16.4	26.4	24.6
P11	7.6	6.9	13.0	9.9	21.7	18.9	40.3	42.8	10.8	11.1	13.6	10.5	11.0	10.0
P12	14.0	9.9	18.6	9.2	36.7	26.2	12.2	11.9	66.6	26.9	–	34.9	–	54.6

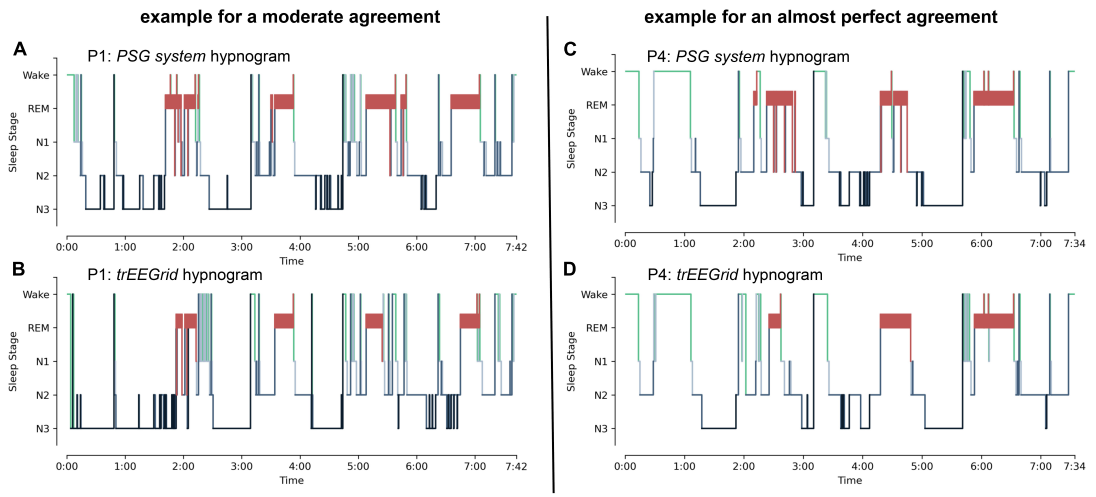


FIGURE 4 | Exemplary hypnograms, shown for participants P1 and P4. **(A,B)** Hypnograms of P1 show a moderate agreement. **(C,D)** Hypnograms of P4 show an almost perfect agreement. The sleep stages are defined according to the AASM as: Art (Artifact), Wake (Wakefulness), REM (Rapid Eye Movement), N1 (non-REM1), N2 (non-REM2) and N3 (non-REM3).

participants). To calculate the median correlation coefficient over epochs or over participants, the single correlation coefficients were first Fisher Z transformed than the median was formed. In a final step, its result was transformed back, by calculating its hyperbolic tangent. The transformation is necessary to approximate the normal distribution of correlation coefficients and reduce the bias of their average (Silver and Dunlap, 1987).

RESULTS

Validation Study

Channel Impedances

Table 2 shows impedance values at 5 min after the start and before the end of the night measurements. It is noticeable, that chin electrodes (R6 and R7) lost connection during the measurement for six out of 12 participants and the electrode in front of the ear (R4) for four out of 12 participants. In 11 out of 15 cases the loose electrodes reconnected overnight. Overall impedances dropped significantly at the end of the night [$F(1,11) = 11.8$; $p < 0.01$]. Across electrodes there was a significant difference of impedance [$F(6,66) = 5.1$; $p < 0.01$], but no significant difference across electrodes and overnight combined [$F(6,66) = 2.3$; $p = 0.08$].

Analysis of Hypnograms

Hypnograms of both systems are shown in **Figure 4** for two participants exemplarily (P1 and P4). The lowest reliability tests result of a single participant was scored for P1 (shown in **Figures 4A,B**) with Cohen's $\kappa = 0.58 \pm 0.02$ (moderate agreement). The highest tests result was scored for P4 (shown

TABLE 3 | Statistical reliability for individual participants (standard error ± 0.02).

Participant	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
Cohen's κ	0.58	0.75	0.77	0.83	0.65	0.60	0.77	0.83	0.70	0.58	0.76	0.64

in **Figures 4C,D**) with Cohen's κ = 0.83 ± 0.01 (almost perfect agreement). All single participant test results are shown in **Table 3**.

The confusion matrix corresponding to hypnograms of all participants and systems combined is shown in **Table 4**. The respective statistical reliability test results over all participants and corresponding standard errors are shown in the following for either every single sleep stage vs. the rest or all sleep stages combined:

Wake: Cohen's κ = 0.75 ± 0.01 (substantial agreement).
N1: Cohen's κ = 0.46 ± 0.02 (moderate agreement).
N2: Cohen's κ = 0.66 ± 0.01 (substantial agreement).
N3: Cohen's κ = 0.79 ± 0.01 (substantial agreement).
REM: Cohen's κ = 0.76 ± 0.01 (substantial agreement).
All: Cohen's κ = 0.70 ± 0.01 (substantial agreement).

Correlation Analysis

Figure 5 (top) shows the results of all three test conditions of the correlation analysis. It should be noted that all annotations of channel R6 and all related channel combinations were discarded for participant 7 as well as channel R7 and all related channel combinations for participant 7, since channel impedance exceeded 60 kΩ at all annotation windows. In all three conditions, clusters of *trEEGrid* channel combinations are visible that have similar absolute correlation coefficients. Overall, R5-R7 and R4-R5 scored the highest absolute correlations to sleep spindles measured at O2_M2 of the *PSG system* (0.15 and −0.14), R1-R7 and R1-R6 the highest to C4_M2 (0.32 and 0.31) and finally R1-R6 and R1 the highest to F4_M2 (0.46 and 0.43). In all three conditions, it is noticeable that sleep spindles measured at O2_M2, C4_M2 and F4_M2 are best represented in EEG of *trEEGrid* channel linear combinations that either point into the direction of the respective *PSG system* electrodes, or are in parallel to its alignment [as shown in **Figure 5** (bottom)]. This result is exemplary shown in **Figure 6**, by comparing a single annotated sleep spindle of participant P4 recorded at C4_M2 of the *PSG system* with the EEG recorded at R5 and R1-R7 of the *trEEGrid*.

These results further confirm the assumption made in da Silva Souto et al. (2021) that combinations of ear EEG channels can be used to estimate the EEG measured at distant locations of the scalp.

Self-Application Study

Figure 7 shows the answers to the self-applicability questionnaire in box plots. Seven out of 12 participants answered the question "How easy was it to put on the chest belt?" with "very much" or "rather." All 12 participants answered the question "How easy was it to prepare the skin?" with "very much" or "rather." 10 out of 12 participants answered the question "How easy was it to find the correct electrode positions?" with "very much" or "rather." Nine of 12 participants answered the question "How easy was it to apply the electrodes at the correct positions?" with "very much" or "rather." 10 out of 12 participants answered the question "How easy was it to connect the electrode grid to the amplifier?" with "very much" or "rather." 10 out of 12 participants answered the question "How confident did you feel during the self-application process?" with "very much" or "rather." All 12 participants answered the question "How confident are you of having applied the electrode grid correctly?" with "very much" or "rather." Asked for the difficulties experienced during the self-application process in a free text format, five participants mentioned the positioning behind the ear as giving them the most difficulty, one mentioned difficulty with using the mirror image, one mentioned the order of application of the different electrodes, and one noted that self-applying alone without assistance was the most difficult part.

Table 5 shows the set-up time of each participant in minutes. On group average, the mean set up time was 12 min (rounded to a 30 s interval) with a range of 8:00 to 19:30 min. The impedance values after 20 min of wear ranged between 5.1 and 98.6 kΩ, excluding values above 100 kΩ. Comparing overall impedances at the start of the overnight measurement of the *validation study* with the recorded impedances of the *self-application study* showed no significant difference [$F(1,11) = 1.3$; $p = 0.28$]. Across electrodes there was a significant difference of impedance [$F(6,66) = 9.1$; $p < 0.01$], and a significant difference across electrodes and between the two measurements combined [$F(6,66) = 3.7$; $p = 0.04$]. Importantly, this interaction is not driven by impedance differences between the two measurements at any electrode, but rather by divergent differences between electrodes within each measurement.

DISCUSSION

In the present study we tested the feasibility of a new sensory grid design (referred to as *trEEGrid*) consisting of nine pre-gelled ECG electrodes around the eye, ear and chin for standard sleep

TABLE 4 | The hypnograms of all participants respective to each system were concatenated to create this confusion matrix.

trEEGrid vs. PSG system	Art	W	N1	N2	N3	REM
Art	0	0	0	0	0	5
W	2	934	110	40	10	15
N1	1	175	476	159	14	32
N2	8	90	204	2,832	345	92
N3	0	15	14	495	2,620	4
REM	0	88	222	219	24	1,387

Each count refers to the annotation of a single 30 s segment.

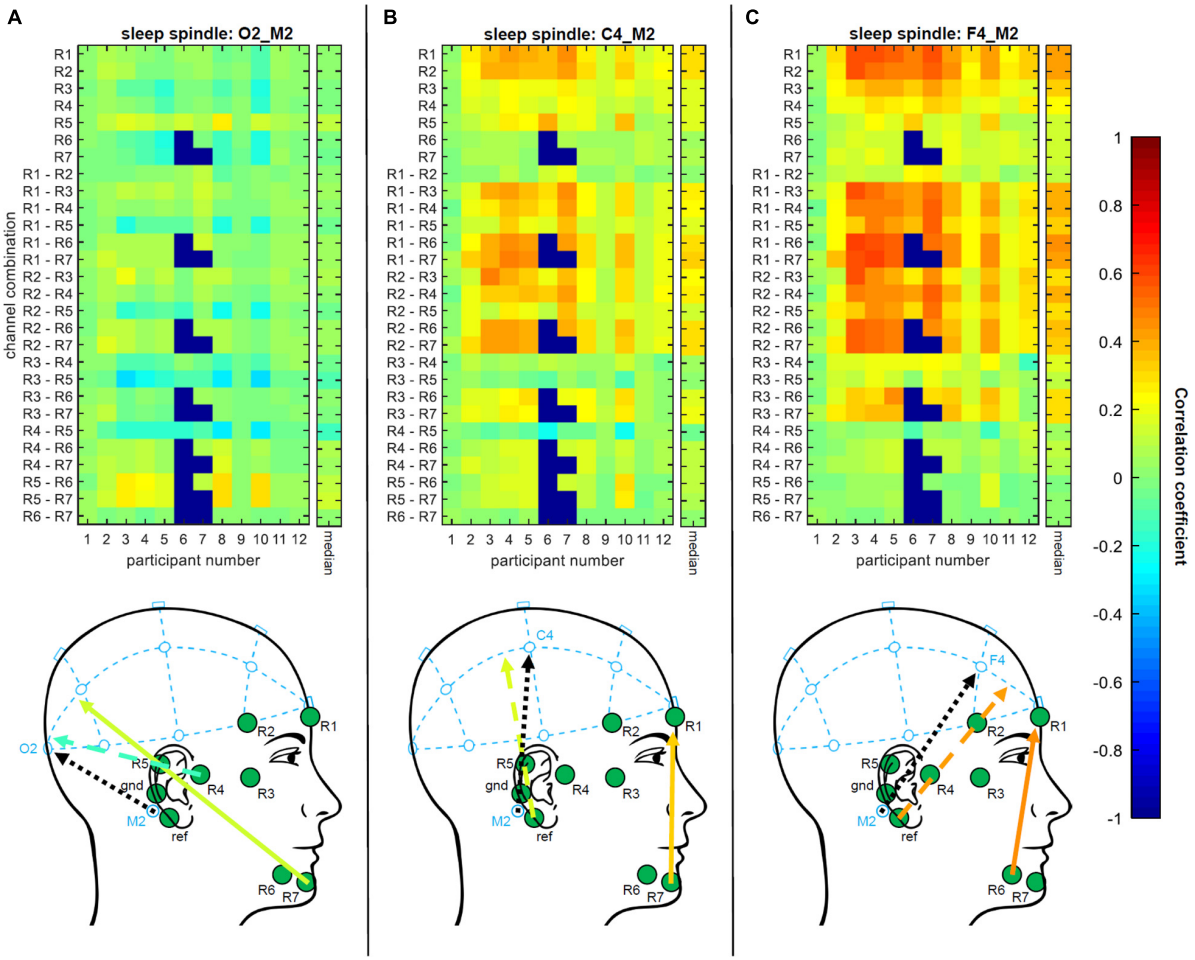


FIGURE 5 | Top: Single participant and median results of all three conditions of the correlation analysis. Correlation coefficients between sleep spindles annotated at EEG of PSG system (using **(A)** O2_M2, **(B)** C4_M2 and **(C)** F4_M2) and the corresponding EEG of every possible single channel and bipolar channel linear combinations of *trEEGrid* electrodes are calculated, Fisher Z transformed, formed the median over epochs and participants, and transformed back. **Bottom:** Visual comparison of alignment of *PSG system* channels (dotted black arrows: O2_M2, C4_M2 and F4_M2) and corresponding *trEEGrid* channel combinations that scored highest median correlation coefficients (solid colored arrows). Feasible linear combinations of *trEEGrid* channels placed around ear are shown for each condition (dashed colored arrows). Colors represent the magnitude of correlations. Based on the correlation analysis *trEEGrid* channel linear combination that either point into the direction of O2_M2, C4_M2 and F4_M2, respectively, or are in parallel to its alignment, result in highest absolute correlation (negative correlations point in the opposing direction of the channel combination) and therefore are best suited to represent sleep spindle reference measurements of the *PSG system* (O2_M2, C4_M2 and F4_M2).

staging recordings. The *trEEGrid* was used in a mobile setting and validated by comparing it to concurrent recordings obtained from a commercial *PSG system*.

Overnight impedance measurements of the *trEEGrid* showed that pre-gelled ECG electrodes were feasible for sleep EEG measurements. Further it was interesting that in 11 out of 15 cases ECG electrodes that lost connection at the start of the measurement recovered connection during the night. This could be a benefit of pre-gelled ECG electrodes over conventional electrodes using electrolyte gels that solidify over time. In this case, recovery of the connection would not be possible after breakage of the solidified gel.

We compared hypnogram annotations of both systems by calculating inter-rater reliability values and found, overall, a substantial agreement between both systems for all participants. Further reliability tests of single sleep stages showed agreements comparable to the literature (Danker-Hopfe et al., 2009), and very similar to previous ear-EEG sleep studies (i.e., Sterr et al., 2018). It must be taken into account that the signals based on the linear combinations of the *trEEGrid* are similar but still different to the signals of the classical EEG electrode positions. This poses a special challenge to the expert scorer, who is used to the classical setting. Automated sleep staging approaches based on machine learning could possibly achieve even better results, in particular when approaches are trained on *trEEGrid* data.

We compared sleep spindles annotated in the *PSG system* with the respective recordings of different *trEEGrid* channel linear combinations in a correlation analysis. In agreement with da Silva Souto et al. (2021), the results provide further evidence that linear combinations of ear EEG channels can be used

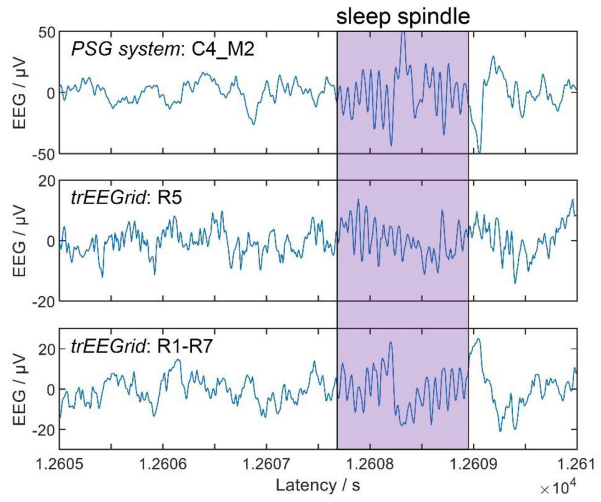


FIGURE 6 | Example showing a single annotated sleep spindle of participant P4, captured by the *PSG system* (**top**, channel C4_M2) and *trEEGrid* channels R5 **(middle)** and R1-R7 **(bottom)**.

TABLE 5 | Results of *self-application study* measurements.

Participants	Setup time [min]	Impedance after 20 min [kΩ]						
		R1	R2	R3	R4	R5	R6	R7
SA01	14:30	26.3	–	23.9	86.5	70.3	15.1	42.9
SA02	11:30	59.2	67.5	30.3	98.6	92.2	35.4	59.3
SA03	08:00	6.5	25.5	23.7	29.7	34.5	35.9	–
SA04	13:00	18.3	44.3	17.4	9.7	52.4	16.7	5.1
SA05	11:00	–	–	–	–	–	–	–
SA06	11:00	8.9	7.1	10.5	43.1	42.6	31.7	–
SA07	14:00	23.2	45.7	36.0	69.1	–	11.9	9.5
SA08	19:30	19.9	16.9	27.6	35.6	–	10.0	5.5
SA09	10:00	19.7	61.4	23.6	–	–	–	–
SA10	09:30	8.5	27.6	13.3	27.8	–	15.6	–
SA11	15:00	18.1	28.2	9.9	–	–	22.1	13.3
SA12	09:30	24.3	73.6	21.4	25.5	91.4	16.0	19.3

For each participant the time needed for self-application of the trEEGrid is given in minutes (rounded to 30 s intervals).
Impedance values measured after 20 min of wear are given in kΩ (impedances above 100 kΩ are marked as "-").

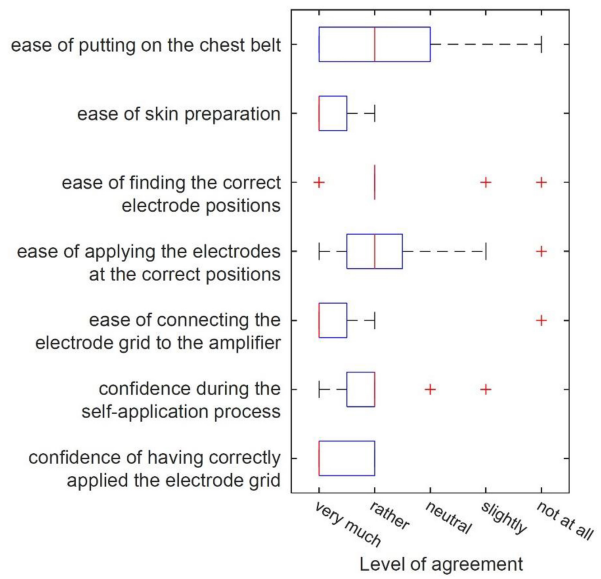


FIGURE 7 | Box plots of answers given by the 12 participants of the *self-application study* concerning different aspects of self-applicability.

to extrapolate information measured at traditional locations on the scalp, like O2 and C4. This motivates the choice of linear combinations of *trEEGrid* channels used for hypnogram annotations. A preliminary test showed that it is possible for the expert scorer to annotate sleep spindles in the EEG recording of the *trEEGrid*. In a future study it would be interesting to analyze if these annotations would differ from those made in a conventional *PSG system*.

We showed that the *trEEGrid* system is self-applicable by healthy participants with only remote instruction. Participants had little trouble finding the correct positions and applied the self-adhesive grid to facial areas adequately. Note that self-application was supervised, so the experimenter was able to give feedback during the process. In a real-life scenario, the goal would be to offer a system that does not rely on an additional person guiding the process. Also, participants in this study were predominantly young adults who had no motor, visual, or hearing impairments that hindered their ability to understand and follow the instructions. For applications that focus on diagnosing or monitoring disorders associated with sleep, older adults may be more afflicted and therefore make up a larger percentage of the user base. In half of the participants, some electrodes lost connection during the measurement and prevented adequate data acquisition. This happened more often with chin electrodes than with any other electrodes. Reasons included movement from eating and talking after applying the grid or insufficient adhesion on a bearded face. Timing the application of the sensors closer to bedtime for a sleep measurement may help to keep sensors in place for a longer period of time, though one participant reported loosening of a chin electrode within 30 min of application.

The overall comparison of impedance values between experimenter-application and self-application showed no significant difference. One major challenge in the application of the *trEEGrid* seemed to be electrode R5, sitting at the top behind the ear. In the *validation study*, this channel lost connection only for one out of 12 participants; in the *self-application study,* however, this channel remained out of range for six out of 12 participants. The position R5 poses a challenge by being both visually occluded and not completely hairless for most participants, therefore rendering it difficult to achieve good signal quality without assistance.

Remarkably, most participants found the electrode grid to be sized exactly right for their faces. Although we did not collect data on facial metrics of the participants, human heads come

in different shapes and sizes. The prototype used in the studies was built of single pre-gelled electrodes fixed into a foam grid for exact placement and additional adhesion. The flexibility of the foam material allowed participants to place sensors where they needed to be while still using the grid shape as a guide. This feature of the grid prototype will be translated into future versions which should be stretchable PCB based as shown in **Figure 1C**, allowing the stretchiness and flexibility of a foam grid while being lighter and smaller.

In the past few years, a variety of new materials for biosignal detection have been introduced which offer fascinating possibilities for face application, including facial tattoos (Shustak et al., 2018), dry film electrodes that rely on surface tension caused by the person's sweat for adhesion (Nawrocki et al., 2018) or conductive aerogel film (Saadatnia et al., 2021). These sensor materials have in common that multiple sensor location can be arranged on a grid which replaces the need for cables and acts as a guide for correct placement at the same time.

In conclusion, the use of self-applicable, comfortable and discrete sensor solutions around the ear could advance future sleep diagnostics by facilitating repeated home-sleep EEG acquisition. Beyond sleep, future disposable flex-printed arrays could be made out of stretchable material, thereby further increasing wearing comfort and signal quality. We argue that improving wearing comfort and unobtrusiveness of sensors is of crucial importance for achieving hassle-free, long-term EEG monitoring solutions, which are of value in health screenings and at-risk neurological and psychiatric patients (Fernandez and Lüthi, 2020; Mander, 2020).

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 Research Ethics Committee, University of Oldenburg, Oldenburg, Germany. The participants provided their written informed consent to participate in the 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

CS, WP, and KW designed the experiments. WP and CS collected the data. MP scored all the sleep data and coded grapho-element annotations. CS and WP analyzed the data. SD provided feedback during all the stages of conducting the studies. WP, CS, SD, and KW wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The Oldenburg Branch for Hearing, Speech and Audio Technology HSA of the Fraunhofer IDMT is funded in the program (Vorab) by the Lower Saxony Ministry of Science and Culture (MWK) and the Volkswagen Foundation for its further development.

ACKNOWLEDGMENTS

We would like to thank all those participating in the study.

REFERENCES

Arnal, P. J., Thorey, V., Ballard, M. E., Hernandez, A. B., Guillot, A., Jourde, H., et al. (2019). The Dreem headband as an alternative to polysomnography for EEG signal acquisition and sleep staging. *bioRxiv [Preprint]*. doi: 10.1093/sleep/zsaa097

Combrisson, E., Vallat, R., Eichenlaub, J. B., O'Reilly, C., Lajnef, T., Guillot, A., et al. (2017). Sleep: an open-source python software for visualization, analysis, and staging of sleep data. *Front. Neuroinform.* 11:60. doi: 10.3389/fninf.2017.00060

da Silva Souto, C. F., Pätzold, W., Wolf, K. I., Paul, M., Matthiesen, I., Bleichner, M. G., et al. (2021). Flex-Printed ear-EEG sensors for adequate sleep staging at home. *Front. Digit. Health* 3:688122. doi: 10.3389/fdgth.2021.688122

Danker-Hopfe, H., Anderer, P., Zeitlhofer, J., Boeck, M., Dorn, H., Gruber, G., et al. (2009). Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard. *J. Sleep Res.* 18, 74–84. doi: 10.1111/j.1365-2869.2008.00700.x

Debener, S., Emkes, R., Vos, M. D., and Bleichner, M. (2015). Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear. *Sci. Rep.* 5:16743. doi: 10.1038/srep16743

Fernandez, L. M., and Lüthi, A. (2020). Sleep spindles: mechanisms and functions. *Physiol. Rev.* 100, 805–868. doi: 10.1152/physrev.00042.2018

Figorilli, M., Marques, A. R., Vidal, T., Delaby, L., Meloni, M., Pereira, B., et al. (2020). Does REM sleep behavior disorder change in the progression of Parkinson's disease? *Sleep Med.* 68, 190–198. doi: 10.1016/j.sleep.2019.12.013

Goldstone, A., Javitz, H. S., Claudatos, S. A., Buysse, D. J., Hasler, B. P., de Zambotti, M., et al. (2020). Sleep disturbance predicts depression symptoms in early adolescence: initial findings from the adolescent brain cognitive development study. *J. Adolesc. Health* 66, 567–574. doi: 10.1016/j.jadohealth.2019.12.005

Jahrami, H., BaHammam, A. S., Bragazzi, N. L., Saif, Z., Faris, M., and Vitiello, M. V. (2021). Sleep problems during the COVID-19 pandemic by population: a systematic review and meta-analysis. *J. Clin. Sleep Med.* 17, 299–313. doi: 10.5664/jcsm.8930

Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174.

Mander, B. A. (2020). Local sleep and Alzheimer's disease pathophysiology. *Front. Neurosci.* 14:525970. doi: 10.3389/fnins.2020.525970

Miettinen, T., Myllymaa, K., Westeren-Punnonen, S., Ahlberg, J., Hukkanen, T., Töyräs, J., et al. (2017). Success rate and technical quality of home polysomnography with self-applicable electrode set in subjects with possible sleep bruxism. *IEEE J. Biomed. Health Inform.* 22, 1124–1132. doi: 10.1109/JBHI.2017.2741522

Miettinen, T., Myllymaa, K., Westeren-Punnonen, S., Ahlberg, J., Hukkanen, T., Toyras, J., et al. (2018). Success rate and technical quality of home polysomnography with self-applicable electrode set in subjects with possible

sleep bruxism. *IEEE J. Biomed. Health Inform.* 22, 1124–1132. doi: 10.1109/JBHI.2017.2741522

Mikkelsen, K. B., Kappel, S. L., Hemmsen, M. C., Rank, M. L., and Kidmose, P. (2019). "Discrimination of sleep spindles in Ear-EEG," in *Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Piscataway, NJ: IEEE), 6697–6700. doi: 10.1109/EMBC.2019.8857114

Nawrocki, R. A., Jin, H., Lee, S., Yokota, T., Sekino, M., and Someya, T. (2018). Self-Adhesive and ultra-conformable, sub-300 nm dry thin-film electrodes for surface monitoring of biopotentials. *Adv. Funct. Mater.* 28:1803279. doi: 10.1002/adfm.201803279

Paunio, T., Korhonen, T., Hublin, C., Partinen, M., Koskenvuo, K., Koskenvuo, M., et al. (2015). Poor sleep predicts symptoms of depression and disability retirement due to depression. *J. Affect. Disord.* 172, 381–389. doi: 10.1016/j.jad.2014.10.002

Saadatnia, Z., GhaffariMosanenzadeh, S., Marquez Chin, M., Naguib, H. E., and Popovic, M. R. (2021). Flexible, air dryable, and fiber modified aerogel-based wet electrode for electrophysiological monitoring. *IEEE Trans. Biomed. Eng.* 68, 1820–1827. doi: 10.1109/tbme.2020.3022615

Shustak, S., Inzelberg, L., Steinberg, S., Rand, D., David-Pur, M., Hillel, I., et al. (2018). Home monitoring of sleep with a temporary-tattoo EEG, EOG and EMG electrode array: a feasibility study. *J. Neural Eng.* 16:026024. doi: 10.1088/1741-2552/aafa05

Silver, N. C., and Dunlap, W. P. (1987). Averaging correlation coefficients: should Fisher's z transformation be used? *J. Appl. Psychol.* 72, 146–148.

Sterr, A., Ebajemito, J. K., Mikkelsen, K. B., Bonmati-Carrion, M. A., Santhi, N., Della Monica, C., et al. (2018). Sleep EEG derived from behind-the-ear electrodes (cEEGrid) compared to standard polysomnography: a proof of concept study. *Front. Hum. Neurosci.* 12:452. doi: 10.3389/fnhum.2018.00452

Tamaki, M., Bang, J. W., Watanabe, T., and Sasaki, Y. (2016). Night watch in one brain hemisphere during sleep associated with the first-night effect in humans. *Curr. Biol.* 26, 1190–1194. doi: 10.1016/j.cub.2016.02.063

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

frontiers | Frontiers in Neuroscience

ORIGINAL RESEARCH
published: 30 June 2022
doi: 10.3389/fnins.2022.873201

Comparing In-ear EOG for Eye-Movement Estimation With Eye-Tracking: Accuracy, Calibration, and Speech Comprehension

Martin A. Skoglund[1,2*], *Martin Andersen*[3], *Martha M. Shiell*[2], *Gitte Keidser*[2,4], *Mike Lind Rank*[3] *and Sergi Rotger-Griful*[2]

[1] Division of Automatic Control, Department of Electrical Engineering, The Institute of Technology, Linköping University, Linkoping, Sweden, [2] Eriksholm Research Centre, Part of Oticon A/S, Snekkersten, Denmark, [3] T&W Engineering A/S, Allerød, Denmark, [4] Department of Behavioral Sciences and Learning, Linneaus Centre Head, Linköping University, Linkoping, Sweden

OPEN ACCESS

Edited by:
Jérémie Voix, École de Technologie Supérieure (ÉTS), Canada

Reviewed by:
Volker Hohmann, University of Oldenburg, Germany
Norbert Dillier, University of Zurich, Switzerland

***Correspondence:**
Martin A. Skoglund
martin.skoglund@liu.se

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

Received: *10 February 2022*
Accepted: *25 May 2022*
Published: *30 June 2022*

Citation:
Skoglund MA, Andersen M, Shiell MM, Keidser G, Rank ML and Rotger-Griful S (2022) Comparing In-ear EOG for Eye-Movement Estimation With Eye-Tracking: Accuracy, Calibration, and Speech Comprehension. Front. Neurosci. 16:873201. doi: 10.3389/fnins.2022.873201

This presentation details and evaluates a method for estimating the attended speaker during a two-person conversation by means of in-ear electro-oculography (EOG). Twenty-five hearing-impaired participants were fitted with molds equipped with EOG electrodes (in-ear EOG) and wore eye-tracking glasses while watching a video of two life-size people in a dialog solving a Diapix task. The dialogue was directionally presented and together with background noise in the frontal hemisphere at 60 dB SPL. During three conditions of steering (none, in-ear EOG, conventional eye-tracking), participants' comprehension was periodically measured using multiple-choice questions. Based on eye movement detection by in-ear EOG or conventional eye-tracking, the estimated attended speaker was amplified by 6 dB. In the in-ear EOG condition, the estimate was based on one selected channel pair of electrodes out of 36 possible electrodes. A novel calibration procedure introducing three different metrics was used to select the measurement channel. The in-ear EOG attended speaker estimates were compared to those of the eye-tracker. Across participants, the mean accuracy of in-ear EOG estimation of the attended speaker was 68%, ranging from 50 to 89%. Based on offline simulation, it was established that higher scoring metrics obtained for a channel with the calibration procedure were significantly associated with better data quality. Results showed a statistically significant improvement in comprehension of about 10% in both steering conditions relative to the no-steering condition. Comprehension in the two steering conditions was not significantly different. Further, better comprehension obtained under the in-ear EOG condition was significantly correlated with more accurate estimation of the attended speaker. In conclusion, this study shows promising results in the use of in-ear EOG for visual attention estimation with potential for applicability in hearing assistive devices.

Keywords: EOG, audio-visual, speech comprehension, eye-tracking, in-ear EEG, hearing impairment

1. INTRODUCTION

There is a strong scientific and commercial trend toward developing more realistic assessment methods for the development of new technology. The main reason is the urge to obtain results that are applicable to real-life scenarios. In hearing sciences, and related fields, this concept is referred to as ecological validity (Keidser et al., 2020). Experiments conducted in less controlled environments that better represent the real world may introduce higher variability and unexpected effects in data, making analysis and interpretation of data more challenging. Sensing technologies can improve scene, situation, context, and intention awareness (Mehra et al., 2020; Slaney et al., 2020), and are promising tools for use in research with less experimental control. In hearing research, there is particularly a growing interest in sensing eye movement that is used as a metric to identify what sound (e.g., speech source) the listener is attending to. This information can be used to identify the best signal processing strategy to apply in hearing devices to optimize hearing in hearing-impaired people. For example, in Best et al. (2017) and Roverud et al. (2017), a highly directional beamformer array was steered by the participants' visual attention using a conventional eye-tracker and multi-speaker conversational stimuli. Their results showed improved performance when the speaker was fixed in a single location, but suggest that it is harder to improve speech intelligibility when the target speech location switches in an unpredictable fashion. Also using a conventional eye-tracker, eye-behavior in dyadic conversation was studied in Hadley et al. (2019), where it was found that increasing noise led to more focus on the speaker's mouth, stressing that eye movement may be a good metric for identifying the speaker attended to in a multi-speaker scenario and thus steering hearing device algorithms. But for real-life applications, the use of conventional eye-tracking devices is inconvenient.

Portable electroencephalography (EEG) is a promising technology that has recently received considerable attention within the fields of, for example, neuroscience and psychology. The low price point, ease-of-use, and portability makes portable EEG systems attractive for integration into sensing platforms that can support experiments in real-life scenarios. EEG systems enable attended sound sources to be decoded from cortical brain responses (O'Sullivan et al., 2014; Fuglsang et al., 2017) with realistic sound stimuli, but these systems typically require several seconds of data and are therefore not yet an option for real-time steering applications. In contrast to selective attention, eye movements can be measured with only temple and forehead electrodes *via* electrooculography (EOG). Given this, it is a good candidate for online applications requiring fast estimates (see e.g., Manabe and Fukumoto, 2006; Favre-Felix et al., 2018; Chen et al., 2019; Belkhiria and Peysakhovich, 2021; Gunawardane et al., 2021; Kastrati et al., 2021). With an across-ear referenced setup, EOG in the transverse plane can be measured with in-ear electrodes. This is particularly of interest for integration in hearing devices (Fiedler et al., 2016) when combined with dry-electrode solutions (Kappel and Kidmose, 2015, 2018; Kappel et al., 2018) that can be used without conductive gel and can easily be used in real-life situations. In Favre-Félix et al. (2017) and Hládek et al. (2018), fixation angles of eye-gaze were estimated in real-time with in-ear EOG, showing great potential for hearing device steering applications. Furthermore, auditory attention estimation (Grimm et al., 2018), using direction-of-arrival and EOG with the purpose of estimating probabilistic sound-source localization, has been evaluated for beamforming in hearing aids. This evaluation demonstrated that EOG can successfully assist in analyzing the soundscape.

In previous work conducted at our laboratory (Favre-Félix et al., 2019), an LED-light was placed on each loudspeaker and was used to indicate which loudspeaker the user should steer their attention to, similar to the setup described in Pomper and Chait (2017). The attended loudspeaker was amplified based on the absolute eye gaze angle in the horizontal plane as estimated using EOG, inertial sensors, and magnetometers. The absolute gaze angle was, however, difficult to estimate. The EOG signal is heavily affected by the noise that is associated with, for example, DC drift and muscle activity, which generates large variability in the results. Other potential drawbacks with that setup were that the participants sometimes scanned the scene to detect when LEDs switched, and the use of LED-lights and lack of other visual cues were not particularly representative of real-life scenarios. Furthermore, the study only used eleven hearing-impaired (HI) participants and speech intelligibility, which was measured using the DAT speech corpus (Bo Nielsen et al., 2014), was not analyzed and reported.

To fully take advantage of sensing technology in future experiments, new assessment methods are needed (Carlile and Keidser, 2020; Lunner et al., 2020). To this end, the increased interest in the more ecologically-valid hearing research outcomes has resulted in more frequent use of audio-visual stimuli and the development of more challenging speech paradigms in hearing research, refer to e.g., Llorach et al. (2018), for an overview of advanced setups introduced for that purpose. The impact of visual cues, and the corresponding eye-gaze behavior, on speech comprehension in complex listening conditions has not yet been fully investigated. For this purpose, an audio-visual test paradigm that targets speech comprehension of a natural dialog has recently been developed at our laboratory (see Cabella, 2021 for an application of this paradigm).

The main objective of the current study was to further develop and validate the application of in-ear EOG for attended speaker estimation in a realistic listening situation. For this, we compared in-ear EOG estimation to that obtained with a conventional eye-tracker. Furthermore, given that in-ear EOG signal quality can vary greatly between electrodes par and over time, we proposed and evaluated a calibration method that presents three metrics for visual inspection to evaluate EOG signal quality in order to extract the best channel from 36 possibilities. The method was assessed against the eye-tracking ground truth reference for estimating the attended speaker (i.e., conventional eye-tracking). Finally, we assessed whether in-ear EOG eye-steering improves speech comprehension for HI participants in this realistic listening situation. Performance with EOG steering was compared to performance with eye-steering *via* a conventional eye-tracking device and no eye-steering.

2. MATERIALS AND METHODS

2.1. Participants

Twenty-seven HI test participants were recruited from the Eriksholm clinic test pool based on the following selection criteria:

- Test participants should not have been previously exposed to audio-visual stimuli.
- Test participants should be native Danish speakers.
- Test participants should not use eye glasses, unless their sight deficit was sufficiently negligible to not impair their vision of the experiment stimulus, or they could replace their eye glasses with contact lenses during testing.
- Test participants should have mild to moderate hearing loss with a maximum of 60 dB at all frequencies. This constraint could be relaxed to 70 dB at the highest frequency i.e.,8 kHz.
- Test participants should not have large hearing-level asymmetries:
  - Asymmetry at each frequency should not exceed 20 dB.
  - Average asymmetry across frequencies should not exceed 10 dB.

All participants signed a consent form and the experiment was approved by The Ethics Committee of the Capital Region in Denmark (H-20030989). Out of the 27 participants, one was excluded due to unavailability to complete all experiment sessions, and one was excluded due to the inability to perform the task. For each of the remaining 25 participants, we collected data for 6 different conditions presented in separate blocks, making a total of 25*6=150 data blocks. Of these data blocks, four were discarded due to different issues, e.g., missing data or missing triggers, leading to a total of 146 valid data blocks. The gender distribution of the final sample was 11 females and 14 males, and the ages were distributed with a mean of $\mu = 69.5$ *years* and a SD of $\sigma = 8$ *years*. Thresholds were measured at each audiometric frequency from 125 to 8,000 Hz. The audiograms of all participants who are included in the results are shown in **Figure 1**.

2.2. Technical Setup

The test was executed at the Eriksholm Research Centre (ERH) in June 2021. Prior to the experiment, the participants had ear impressions made, from which silicone molds with embedded in-ear electrodes were produced, as shown in **Figure 2** for an example. The technical setup of the experiment is illustrated in **Figure 3**. Participants were seated in a comfortable chair in front of an 88 inch curved TV. Participants were fitted with EOG ear molds made from individual ear impressions equipped with 6 dry contact electrodes, which was connected to a 32 channel Mobita EEG amplifier from TMSi, which sampled at 500 Hz as described in Kappel et al. (2018); Tobii Pro Glasses 2 mobile eye-tracking device, which sampled at 50 Hz, and included reflective markers that were recorded *via* a Vicon motion capture system, which sampled at 100 Hz; and two behind-the-ear (BTE) devices for recording audio, accelerations, rotational velocity, and magnetic field. The BTE data was not used in the current study. The Vicon motion tracker was used to monitor potential movement of the test participants' heads. MATLAB 2019a was used to present the audio-visual stimuli, see Section 2.3, on the TV. Target speech was directionally presented from two loudspeakers, and babble noise in the frontal hemisphere was presented from eight loudspeakers, further explained in Section 2.3, and simultaneously presented from 10 loudspeakers situated in front of the participant *via* both a Fireface UCX soundcard from RME and a Hammerfall DSP Multiface II. The two loudspeakers in green in **Figure 3** presented the target speech and were roughly spatially aligned with the position of the speakers on the screen. For the comprehension task, questions and response options were shown on the screen, and participants answered the questions using a Bluetooth keyboard. Data from all capturing devices (Tobii, EOG, and Vicon) were synchronized with a signal delivered through the sound card. The one-point calibration, as part of the Tobii glasses setup, was done before each recording block if it was deemed necessary. A desktop computer and a laptop were used in this complex setup. The desktop computer recorded Tobii and Vicon data and executed the stimulus MATLAB script. The EOG data was collected on the laptop which also executed the eye-steering algorithms using MATLAB 2016b. To enable communication of the attended speaker between the laptop and the desktop computer, a Maya USB+ 44 soundcard sent audio signals between the computers. **Figure 3** illustrates the test set-up while **Figure 4** describes the test flow.

The hearing loss of the test participants was compensated on the stimulus side based on the audiogram for a better ear. The compensation was computed with the CAMEQ formula for linear hearing aids (Moore and Glasberg, 1998), with the CAMEQ output extrapolated from 6 to 9 kHz using a cubic spline. This way of compensating for hearing loss can lead to a loud environment and that is why the hearing loss of the test participants was restricted to the mild-to-moderate range. The experimenter, who was present in the room during testing, wore ear protection.

2.3. Audio-Visual Stimuli Targeting Speech Comprehension

The stimuli and tasks were derived from Cabella (2021) and consisted of in-house HD video recordings of two speakers engaged in an unscripted conversation in Danish, refer to the example screenshot in **Figure 5**. Speakers were fitted with hands-free microphones and their speech was recorded on two separate audio channels. Two different sets of speakers were used, with pairs have taken either from four paid actors, herein referred to as the ACT material or four volunteers recruited among Eriksholm Research Centre staff, herein referred to as the ERH material. The speakers' conversation consisted of them solving a Diapix task (Baker and Hazan, 2011) wherein they verbally compared two similar drawings to find differences between them. The conversations were clipped into trials ranging between 10 and 39 s long, with each clip ending after the speakers uncovered a difference. These clips were preceded by a central fixation cross on a black screen for 3 s. The ACT material was used to generate a babble-noise that was subsequently used with both materials.

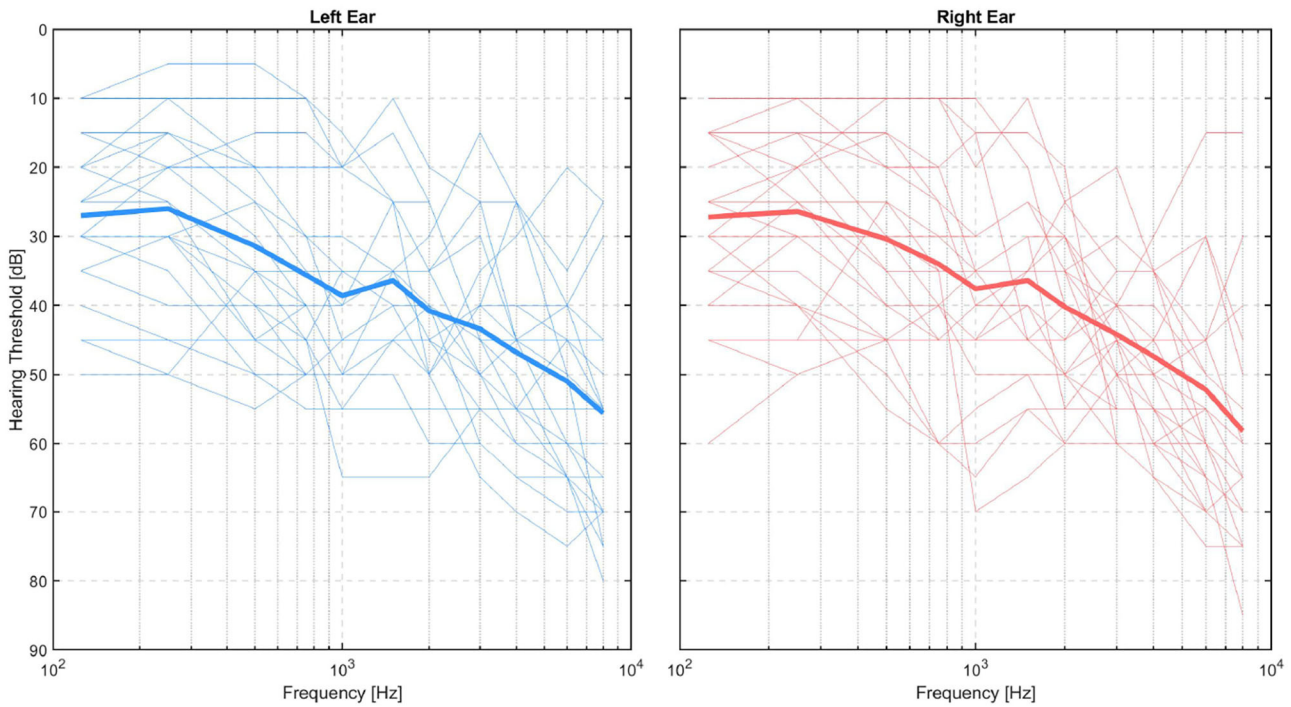


FIGURE 1 | Audiograms of test persons shown separately for left (blue lines) and right (red lines) ears, with the group mean per ear represented by thick lines.

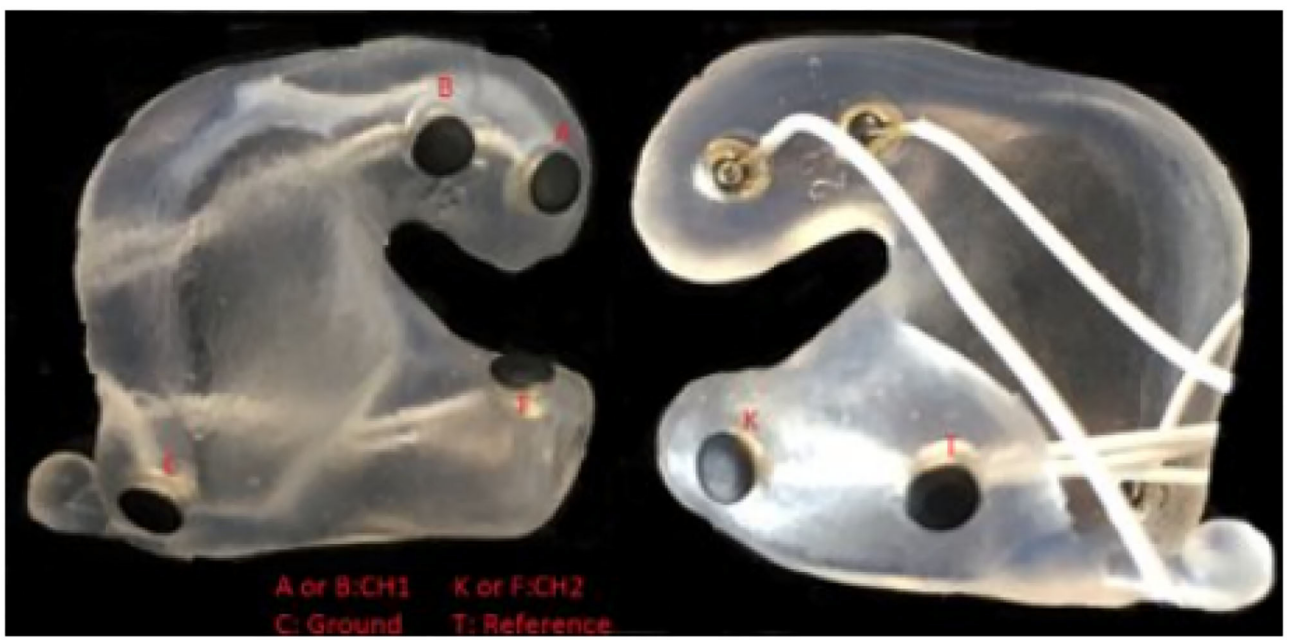


FIGURE 2 | Example of a soft silicone mold equipped with EOG electrodes.

In order to assess how well the listener understood the conversation that was presented in the stimulus, a multiple-choice question with three response options followed each clip. In this question, the participant was asked to identify what the speakers identified as the difference between their drawings. The options consisted of broad categories that described reoccurring differences in the test materials, such that individual response options could be used for multiple different

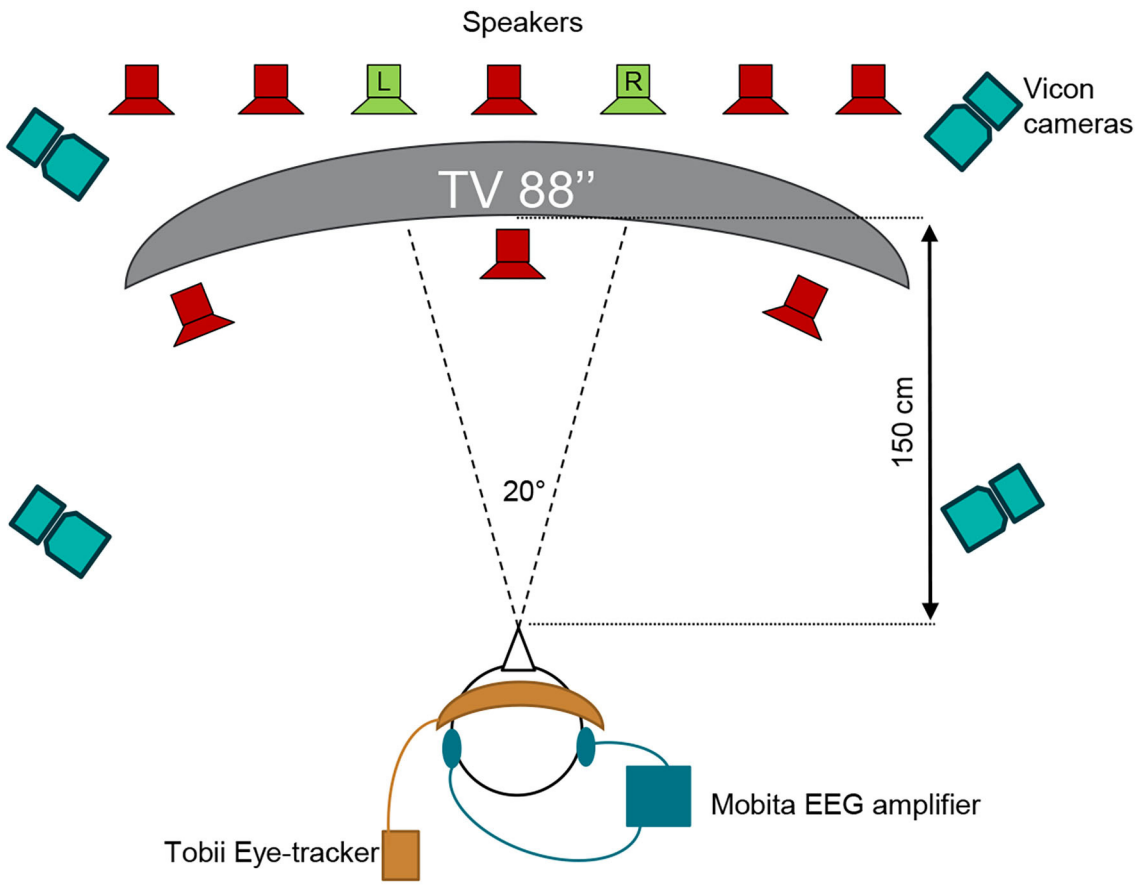


FIGURE 3 | Schematic overview of the test setup with Vicon motion capture cameras, Tobii Pro eye-tracking Glasses 2 used to measure gaze, the Mobita EEG amplifier used to record in-ear EOG signals, and the 88 inches 4K TV used to present life-sized persons in the video stimuli. The two green loudspeakers present the conversation material and are spatially aligned with the two speakers on the TV while the eight red loudspeakers present babble noise.

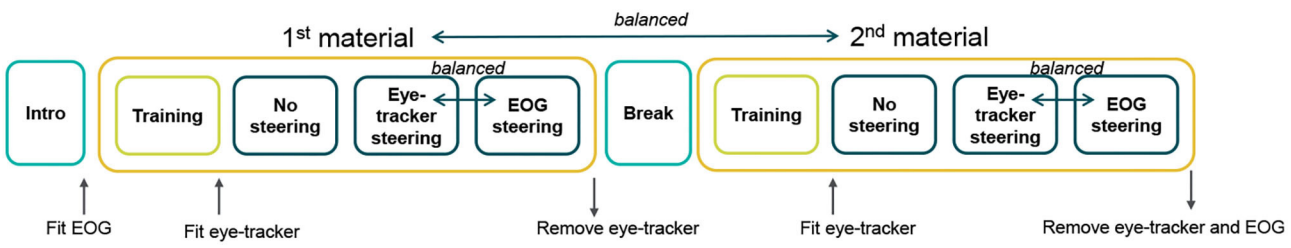


FIGURE 4 | Schematic overview of the procedure used for each test person.

trials. Because these response options were not identical to the words spoken in the conversation, this testing strategy pushed the listener to understand the meaning of the conversation rather than merely recognize words. Note that for the ACT materials, the questions and stimuli were previously used in an experiment with 11 HI listeners (Cabella, 2021). For the ERH materials, the questions were piloted with a small number of normal-hearing and hearing-impaired participants prior to the experiment.

A total of 97 different trials were used in the experiment, with 67 and 30 from each of the ACT and ERH materials, respectively. The ACT material was divided into a practice block of 13 trials, used to familiarize the participants with the task, and 3 blocks of 18 trials each. The ERH material was divided into a practice block of 6 trials, and 3 blocks of 8 trials. Each condition was tested with a separate block of trials from each material, and the same block was used for every participant. This allowed for comparison between participants within a block, for

FIGURE 5 | Still photo from the audio-visual stimuli with two actors in dialog solving a Diapix task (Baker and Hazan, 2011).

the purpose of assessing the success of the proposed calibration procedure. Since the no-steering condition is the most difficult for the participants, it was always presented first for each type of material, thereby reducing the probability of participants giving up from fatigue. This choice was considered in favor of complete block randomization with the assumption that further learning effects after the initial practice blocks are negligible. The trial order was randomized within each condition for each participant, and the order of the steering conditions was counterbalanced between participants.

2.4. Steering Conditions

Each participant carried out the comprehension task under three different conditions:

- No-support: Constant SNR set to 0 dB for the full experiment. Sound levels for the attended speaker, unattended speaker, and noise were all set to 60 dB SPL.
- EOG steering: Sound levels were the same as for the no-support condition, but a 6 dB gain was provided to the estimated attended speaker computed from EOG data.
- Eye-tracker steering: Sound levels were the same as for the no-support condition, but a 6 dB gain was provided to the estimated attended speaker computed from eye tracking data.

Before testing a material condition, the participants completed a training block to familiarize themselves with the task and stimuli. In the training block, the SNR was 10 dB, with the target speech presented at 60 dB SPL and the babble noise at 50 dB SPL.

2.5. Eye-Tracker Reference and Steering

Eye-tracker gaze data in the horizontal azimuth was used to determine which of the two speakers the participant attended at each time point, see yellow asterisk (left speaker) and green asterisk (right speaker) in **Figure 6** for an example. For this, attention to a speaker was defined as gaze within that speaker's respective hemifield. Subsequently, these data were used as a

ground-truth reference for performance in comparison to the EOG steering.

2.6. EOG Signals and Steering

Electro-oculography signals are proportional to eye deflection and are produced by the cornea-retinal potential in the human eye. Here, the potentials were measured with across-ear referenced, dry contact, and in-ear electrodes and were most sensitive to eye-movements in the horizontal azimuth. The dry contact electrodes, also used in Favre-Félix et al. (2019) and described in e.g., Kappel and Kidmose (2015), Kappel and Kidmose (2018), and Kappel et al. (2018), were based on a titanium substrate coated with iridium dioxide and mechanically designed to be embedded into a soft earpiece. There were 6 electrodes per ear and since the Mobita EEG amplifier uses a common mode reference, there were a total of 36 single different possible channel combinations.

2.6.1. Saccade Detection and Attended Speaker Estimation

The electro-oculography steering algorithm estimated which of the two speakers the user attended at any point in time. Due to the skin-electrode interface (Huigen et al., 2002), the signals were vulnerable to low-frequency drift in the same order as the EOG potential, which challenges the calculation of an absolute gaze angle. This drift may be caused by the pre-amplifier, the contact potential changes, or changes in the EOG potential. However, saccades (fast eye-movements), such as those that a listener makes when switching their gaze between the two speakers, generate high-frequency responses that can be distinguished from the drift. We worked with the assumption that saccades were only generated by a switch in the listener's gaze between the speakers and not a switch elsewhere in the scene, and furthermore that the eyes remained fixated (on a speaker) when not performing a saccade. Hence, attended speaker estimation hinged upon accurate and robust saccade detection, which we based on the derivative of the EOG signal. Prior to obtaining the derivative, a second-order Butterworth filter with passband 0.1–1 Hz was used to remove part of the baseline drift and most of the electronic high-frequency noise, such as EEG-based measurement, while keeping sufficient saccade information for analysis. Saccade detection required the signal to conform to the following criteria:

- The derivative of the signal was bounded by a lower and an upper threshold of 0.1 mVs^{-} and 10 mVs^{-}, respectively.
- The duration of a saccade was bounded by a lower and an upper duration threshold of 0.2 and 0.9 s, respectively.
- The magnitude of a saccade was calculated based on the absolute amplitude change, and absolute amplitude changes smaller than 15 mV were removed.
- Saccades with the wrong direction, e.g., a saccade to the left when the left speaker is already attended, were excluded.

The criteria listed above were refined during extensive testing with this specific setup in order to optimize attended speaker selection while mitigating disturbances. Note that with this procedure, there was no need for an absolute mapping between EOG and horizontal gaze location, as only the hemifield separation was considered, i.e., left or right relative to the test person.

2.7. Pre-Block EOG Channel Selection

To select the best of 36 possible channels for estimating the attended speaker, a calibration method was developed. Before each block, a calibration sequence was used consisting of a red dot that the participant was instructed to follow with their gaze. The dot moved between two horizontal positions on the screen in a pre-determined sequence. An example output of the calibration, as visible to the experimenter, can be seen in **Figure 7**. The positions of the dot approximated the speakers' locations at $\pm 10^{\circ}$ in the visual material. This allows various metrics to be computed based on all channel combinations, which were then used to select the best channel. The metrics computed were as follows:

- The correlation between the EOG and the reference signals. The higher the positive correlation, the better the channel.
- The proportion of correctly detected saccades using the EOG steering algorithm. The target value is 100%, corresponding to 6 out of 6, where the first saccade at 5 s is small, see **Figure 7**.
- The saccade-to-Fixation-Ratio (SFR) is the average saccade amplitude divided by the average standard deviation of the fixation. The higher the SFR, the better the channel.

The rationale for the SFR metric was that a high saccade amplitude corresponds to a better channel and that the signal should not fluctuate substantially during fixations, as it is expected to be a stationary mode. Note that SFR is also sufficiently simple to compute online and could potentially be used to monitor signal quality. The calibration output, as exemplified in **Figure 7**, was visually inspected by the experimenter to judge which channels to use. The two experimenters reported that using a combination of the plots and the metrics to select channels provided good support for finding channels and/or spotting errors.

3. RESULTS

3.1. Attended Speaker Accuracy

The steering algorithm computed the attended speaker based on in-ear EOG as a time series, see the example in **Figure 6**, shown by black and magenta dots representing left and right speakers, respectively. The accuracy metric computes which percentage of time the in-ear EOG-estimated attended speaker matched the eye-tracker reference. A value of 100% means that the EOG data matched eye-tracker data perfectly, while the chance level is 50%. Note that the in-ear EOG attended speaker is initialized in the algorithm such that the first seconds of the trial may not reflect algorithm performance but chance, hence the first 3 s were removed in the accuracy measure for each trial. A scoring function was used to evaluate the in-ear EOG attended speaker accuracy based on the eye-tracker-attended speaker. Eye-tracker data within 2° with respect to zero azimuth was considered to be inconclusive and was omitted from the score. The attended speaker accuracy from the experiments across all conditions and materials for all participants is shown in **Figure 8**.

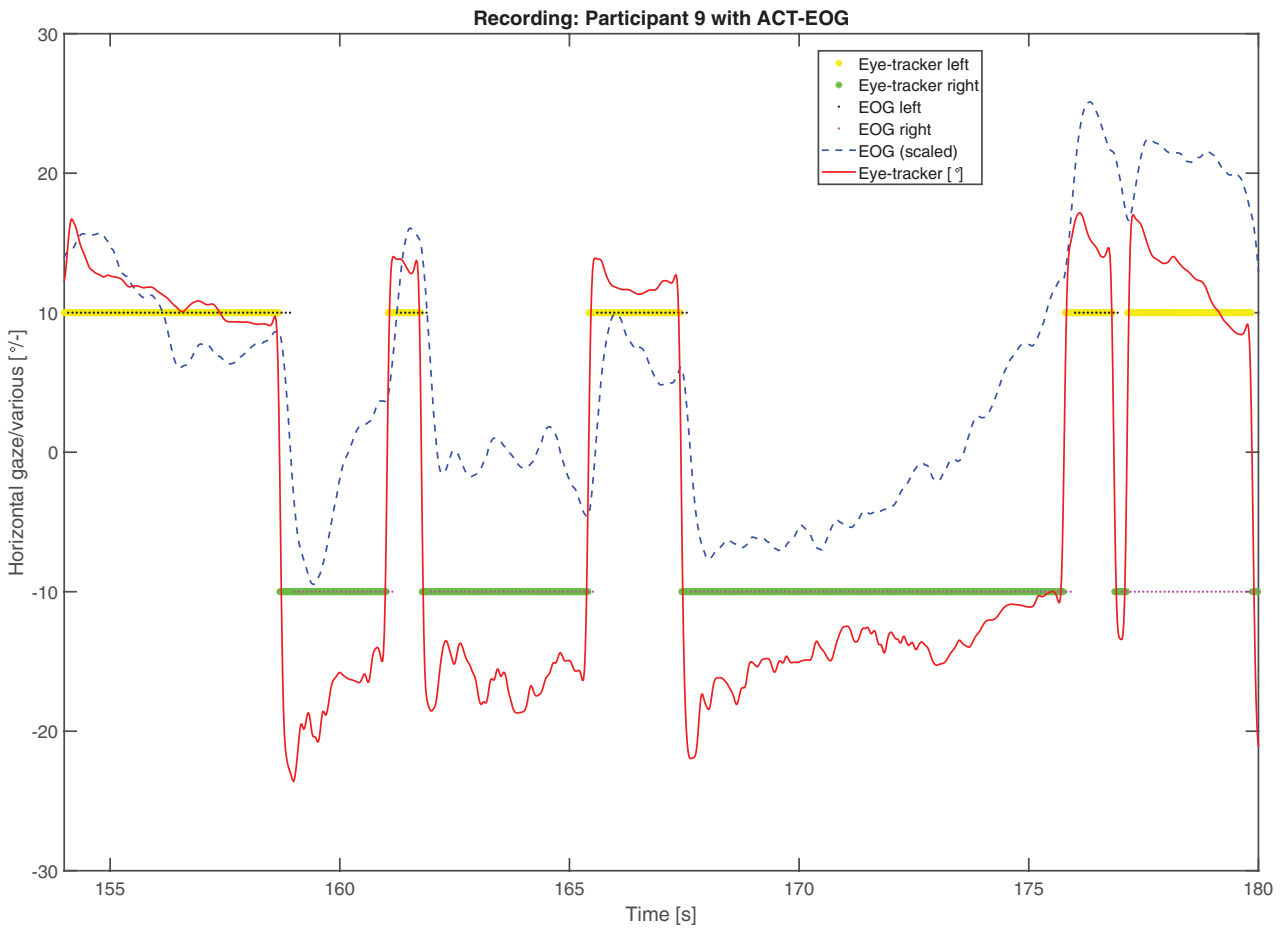


FIGURE 6 | An example of four channels showing the highest correlation (Corr). Reference signal in solid red, EOG signal in green, assumed saccade events in red overlaid on the green line, and detected saccades represented by black dots.

One participant out of the 25 was omitted from this analysis as it had missing data in one or more conditions. The mean accuracy was 68%, the highest accuracy was almost 90%, and the worst was nearly down to chance level. In **Figure 9**, the estimated attended speaker accuracy is shown per condition and material, averaged across all participants. The accuracies per condition and material passed a Lilliefors test of normality. A one-way ANOVA did not support that there were any differences in average attended speaker accuracy between any of the six condition blocks [$F_{(5,\ 126)} = 0.37$, $p = 0.87$].

3.2. Calibration Evaluation

Since the calibration procedure code was developed for both online and offline use, it was possible to calculate the accuracies for all 36 electrode combinations using the recorded data. This allowed for an evaluation of the different strategies of selecting electrodes and assessing selection effectiveness during the experiment. The simulated attended speaker accuracy was computed using four different scenarios that may lead to the selection of different EOG channels. The scenarios were as follows:

- True accuracy: The channels that were selected by the experimenter using the calibration procedure.
- Best accuracy: The channel resulting in the highest accuracy among all 36 without considering the calibration.
- Corr accuracy: The best electrode combination out of all 36 as suggested by ranking the correlation estimated by the calibration procedure prior to each block.
- SFR accuracy: The best electrode combination out of all 36 as suggested by ranking the SFR score estimated by calibration procedure prior to each block.

Due to the technical setup and uncertainties in sampling rates, simulations with the recorded data did not align with the experimental data. For the purpose of this analysis, however, this was not of concern. Rather, it is the relative difference in attended speaker accuracy based on channel selection for the different scenarios that are of interest. The simulated True accuracy gave a mean of 63% compared to the accuracy mean from the online computation at 68%, resulting in a 5% difference in mean accuracy. The other three methods (Best, Corr, and SFR) produced means of 69, 59, and 62%, respectively.

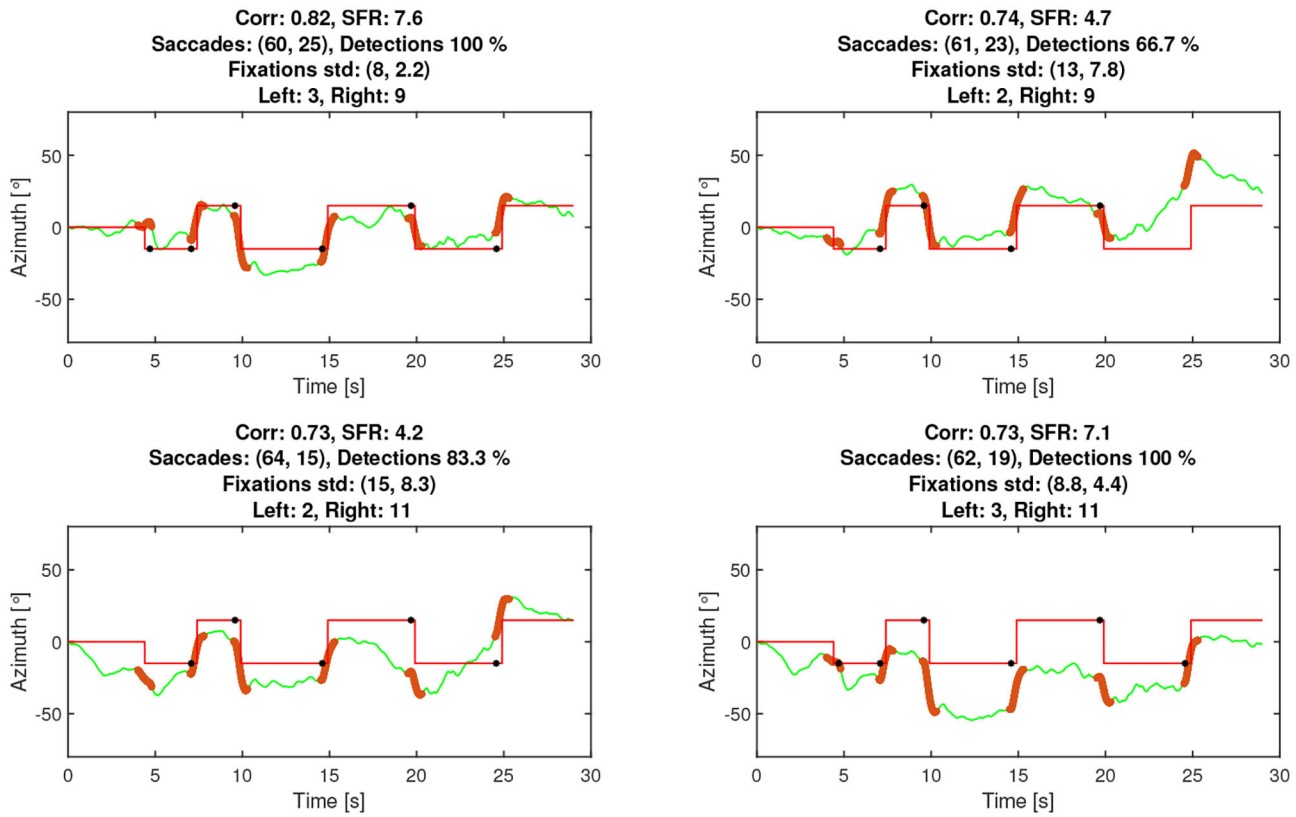


FIGURE 7 | Example of attended speaker estimation data from a trial with participant TP1743. The solid red line represents gaze in degrees from the eye-tracker, the dashed blue line is the in-ear EOG signal, which is scaled for illustration purposes, the yellow and green asterisks are the eye-tracker-attended speakers while the black and magenta small dots are the in-ear EOG-attended speakers. Values below and above 0 deg. refer to left- and right-hand speakers, respectively.

On this basis, we conclude that the experimenters selected effective channels with guidance from the calibration metrics. In **Figure 10**, the three calibration metrics, namely correlation, SFR, and detection (number of agreeable saccades), are plotted vs. the attended speaker accuracy per test participant, averaged across the conditions. All three metrics were related to attended speaker accuracy with correlations over 0.6. In the bottom right-hand plot, the average comprehension scores obtained in the in-ear EOG conditions per participant are shown. These were also positively correlated with attended speaker accuracy ($\rho = 0.45236, p = 0.026459$), such that comprehension scores improved as the EOG steering became more similar to the eye-tracker steering.

3.3. Speech Comprehension

The overall performance scores in terms of speech comprehension in the audio-visual task are shown in **Figure 11**. As the task was a multiple-choice task with 3 possible answers, the data is binomial with the probability of success for a single trial being $\frac{1}{3}$. Based on the cumulative binomial distribution, the comprehension score chance level for the individual participant is 50% for the ERH material and 63% ACT material. Based on the work by Cabella (2021), that tested 7 hearing-impaired participants, the presentation SNR of 0 dB was expected to produce an average score of 60% for the no-support condition, which is sufficiently low to allow for potential improvement with the two steering conditions. As seen in **Figure 6**, this prediction was close to the actual outcome, with 66.7 and 65.3% correct for the ACT and ERH materials, respectively. The mean scores in the ACT-eye-tracker and ACT-EOG conditions were 76.8 and 76.2%, respectively. In **Figure 12**, the individual performance scores are shown together with the mean scores.

The following statistical analysis aimed to evaluate if there was a significant difference between the three steering conditions. One participant was omitted from this analysis as there were missing data in one condition, leaving 24 participants. A Lilliefors tests on each steering condition for each material show that not all data were normally distributed. Since only the steering conditions (no-support, eye-tracker, and EOG) were of interest here, and not the two types of materials, the comprehension scores were averaged across material types. The averaging of the scores was justified by a Mann-Whitney U-test showing no evidence of a significant difference in comprehension score between the two materials in each of the three test conditions: no-support ($U = 597$, $p = 0.86$), eye-tracker ($U = 573$, $p = 0.76$), and EOG ($U = 528$, $p = 0.21$). A Friedman test was conducted on the comprehension scores using the three conditions as independent variables. We found a significant

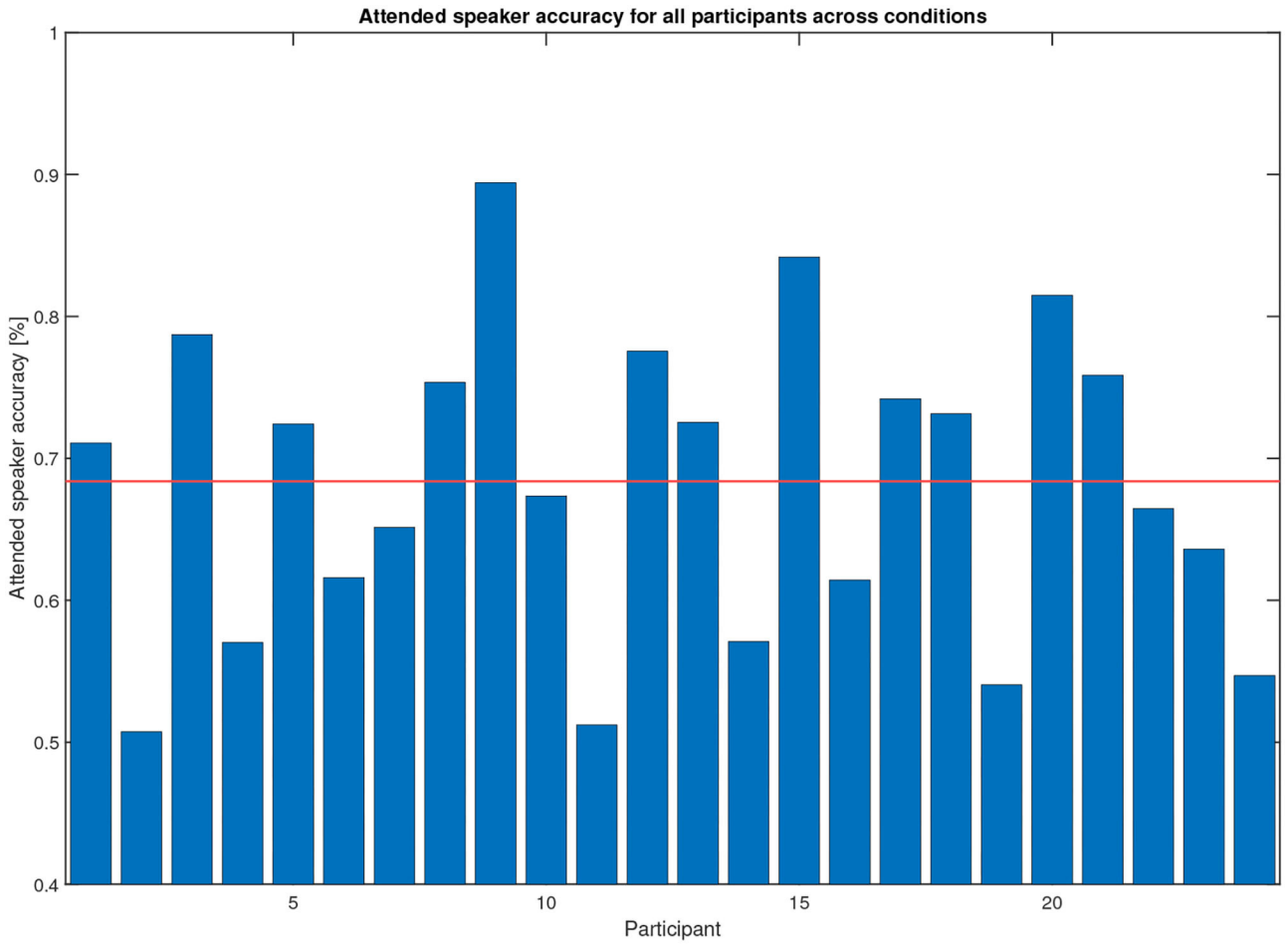


FIGURE 8 | Attended speaker true accuracy across all conditions for the 25 participants (only 24 shown due to missing reference data for one participant). The mean of the true accuracy across participants is 68%. These data were computed during the experiment.

difference in the comprehension scores between the different conditions ($p = 3.5 \times 10^{-5}$, $Q = 20.6$). A *post-hoc* comparison of the comprehension scores in each condition was done with a Tukey-Kramer critical value test calculated at 5%. We found a significant difference between the no-support and both eye-tracker and EOG, with condition differences for no-support and eye-tracker (-0.88, $p = 0.0056$), no-support and EOG (-1.25, $p = 2.89 \times 10^{-5}$). The condition difference between eye-tracker and EOG was not significant at (-0.38, $p = 0.28$). The ranks of eye-tracker and EOG conditions (2.17, 2.54) were both greater than no-support (1.29), supporting that the steering conditions have a positive effect on comprehension.

4. DISCUSSION

As mentioned in the introduction, in-ear EOG has great potential for application in hearing devices to identify relevant speech among multiple speakers that the user wishes to attend to. The experimental setup used here, presenting an audio-visual dialogue, was intentionally limited in order to achieve robust and interpretable data and therefore does not represent the multitude of listening scenarios that hearing device users may be exposed to in everyday life. For example, contrary to the question-answer paradigm used in Best et al. (2017) and Roverud et al. (2017), where several speakers and directions were involved, the turn-taking in the current test, with only two speakers, is predictable. However, listening to dialog is a common real-life situation that can be seen as a building block for more complex multi-party conversations, and findings from this study should be helpful in refining steering technology for investigation in more complex settings.

Most of the participants had reasonable accuracies in the attended speaker EOG estimation as compared to the eye-tracker reference, and the overall mean was 68%. The accuracy of the attended talker estimation based on EOG was not significantly different across conditions or materials, and such a dependency was not expected either. The attended talker estimation based on saccade detection in EOG was robust and generalized well to the participants without the need for individualized parameters, which we believe is key for future applications. Beyond the

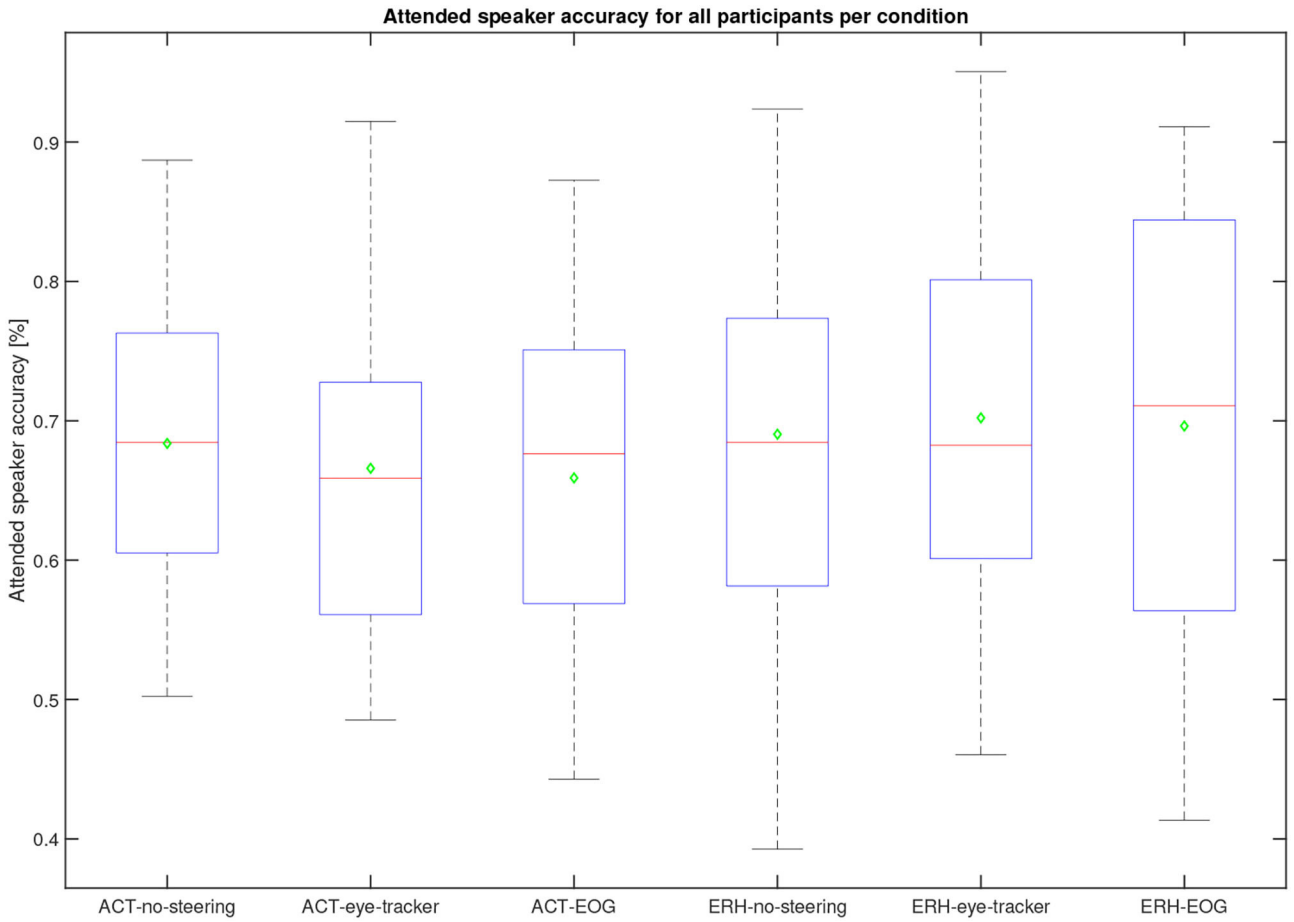


FIGURE 9 | Average speaker accuracy across 22 participants (two were omitted due to missing data for one condition or more) for the different conditions. The median is the red line, the mean is a green diamond, the box represents the 25 and 75*th* percentile, the whiskers extend to the extreme values, and the red crosses represent individual outliers. These data were computed during the experiment. No significant difference is found between any of the conditions and materials.

scope of this paper but of interest for future endeavors on eye-movement analysis with EOG are comparisons of saccade detector designs and their performance; evaluation of other metrics, such as saccade rate and average fixation time with respect to SNR; and absolute angle estimation.

In line with previous work (Huigen et al., 2002; Hládek et al., 2018; Kappel et al., 2018; Favre-Félix et al., 2019), it was observed that some people have mostly distortion free EOG signals, leading to reliable saccade detections, while it can be extremely hard to find distortion free signals for others. Dependent on how well the ear-mold fits in the ear of the participant, a poor electrode-skin interface (Kappel et al., 2018) with weak signals and distortions may result. To support the experimenter in ensuring acceptable signals were obtained, a pre-block calibration procedure was developed utilizing three calibration metrics indicative of channel quality. The calibration procedure supported the finding of the most distortion free, but not necessarily the best, EOG signal to use in the experiments. All three calibration metrics were reasonable indicatives of attended speaker accuracy, with particularly the SFR calculation showing potential as a candidate for automated channel selection. This is because SFR is easily adapted for online signal quality monitoring without the need for reference data as saccades and fixations are already estimated in the algorithm. A restriction with the current setup is the across-ear referenced EOG which requires a wired connection between the ears, severely limiting hearing aid integration. Thus, future work should aim to introduce single ear EOG, which likely has worse SNR, further stressing the importance of calibration and signal quality monitoring procedure. For EOG methods relying on absolute angles, see e.g., Hládek et al. (2018) and Favre-Félix et al. (2019), the calibration is used to map the angles to voltage levels Manabe et al. (2013). These may vary between participants and require a reference for calibration. A potential alternative for hearing aid applications is to estimate speaker directions with binaural direction-of-arrival, see e.g., Braun et al. (2015), Zohourian et al. (2018), and Grimm et al. (2018), which is then used to calibrate the EOG.

Two steering conditions, one using conventional eye-tracking and the other in-ear EOG, provided significant comprehension improvement with reference to the no-steering condition. As

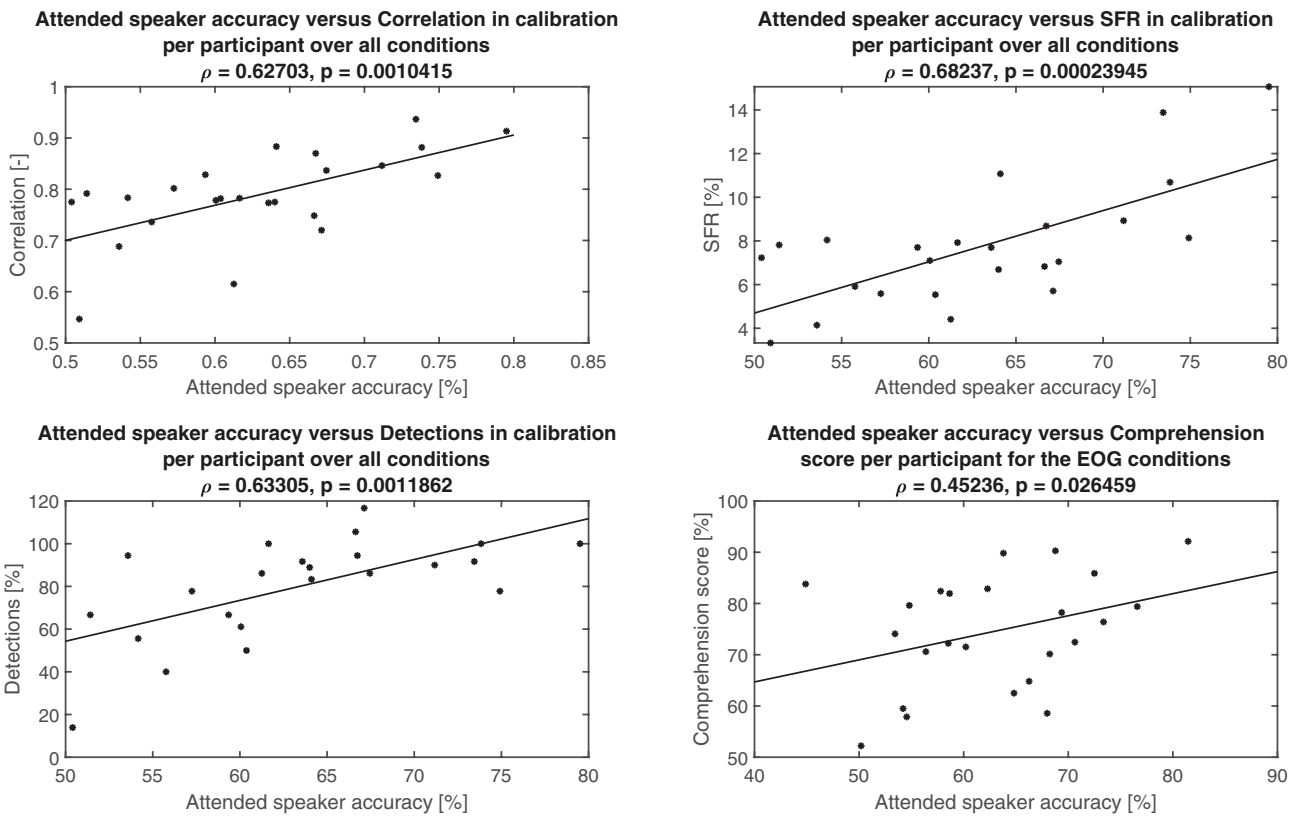


FIGURE 10 | Attended speaker accuracy vs. various metrics per participant together with the least-squares line fit in solid, Pearson's correlation (ρ), and the corresponding p-value. Top left: Correlation metric from calibration procedure. Top right: SFR metric from the calibration procedure. Bottom left: Detection metric from calibration procedure. Note that one outlier showing unreasonable 300% detections was removed. Bottom right: Comprehension score in the EOG conditions.

long as participants paid attention to the active speaker in the dialog, an overall performance improvement in the steering conditions was expected due to the increase in SNR applied to the attended speaker. However, the improvement experienced with such steering systems will depend on several factors, such as the accuracy of attended speaker estimation, the complexity of the stimuli, and the signal processing applied to the attended sound. From previous analyses done in experiments that used the same material, more variation was found in the trials themselves than at different SNR levels (Cabella, 2021). Trials that were considered too easy or too difficult were only used for practice runs, and not used in the experiment. Even so, the remaining material variability resulted in difficulties finding an SNR associated with a 60% comprehension score for all participants in the no-support conditions. Part of this variability may be attributed to the different degrees of hearing loss presented across participants, even if it was compensated for. Ideally, the material variability would have a small influence on the average scores provided a sufficiently large number of participants were tested. Since, for logistic reasons, recruiting a large pool of participants was not possible for this study, it was instead decided to minimize the between-participant variance, and hence the variance within the test conditions, by fixating the trial material for each condition. There is therefore a confound on the material condition difficulty level that cannot be resolved in the analysis. Therefore, the comprehension results should be interpreted with care.

Although there was no significant difference between comprehension performance on the ACT and ERH material within each test condition, a slightly better overall performance was observed in the ERH material than in the ACT material, see **Figure 11**. There are two likely reasons for this. One is that the Eriksholm staff spoke more slowly and clearly than the actors, making the dialog easier to follow, and the other is that the questions developed for the ERH material were simpler and thus required less cognitive effort to answer. The babble noise which was used in both materials was generated from ACT material, which in general was a bit quieter than the ERH material. This means that for the same SNR condition (e.g., ACT-no-support and ERH-no-support), the SNR of the ERH material was slighly 0.83dB A higher than for the ACT material. This was known and accepted prior to the start of the data gathering. It remains for the future to better understand the differences and similarities between the materials, and also between individual trials, and then more systematically assess the effects they have on comprehension scores.

There was no significant difference between performance in the two steering conditions. It had been expected that the

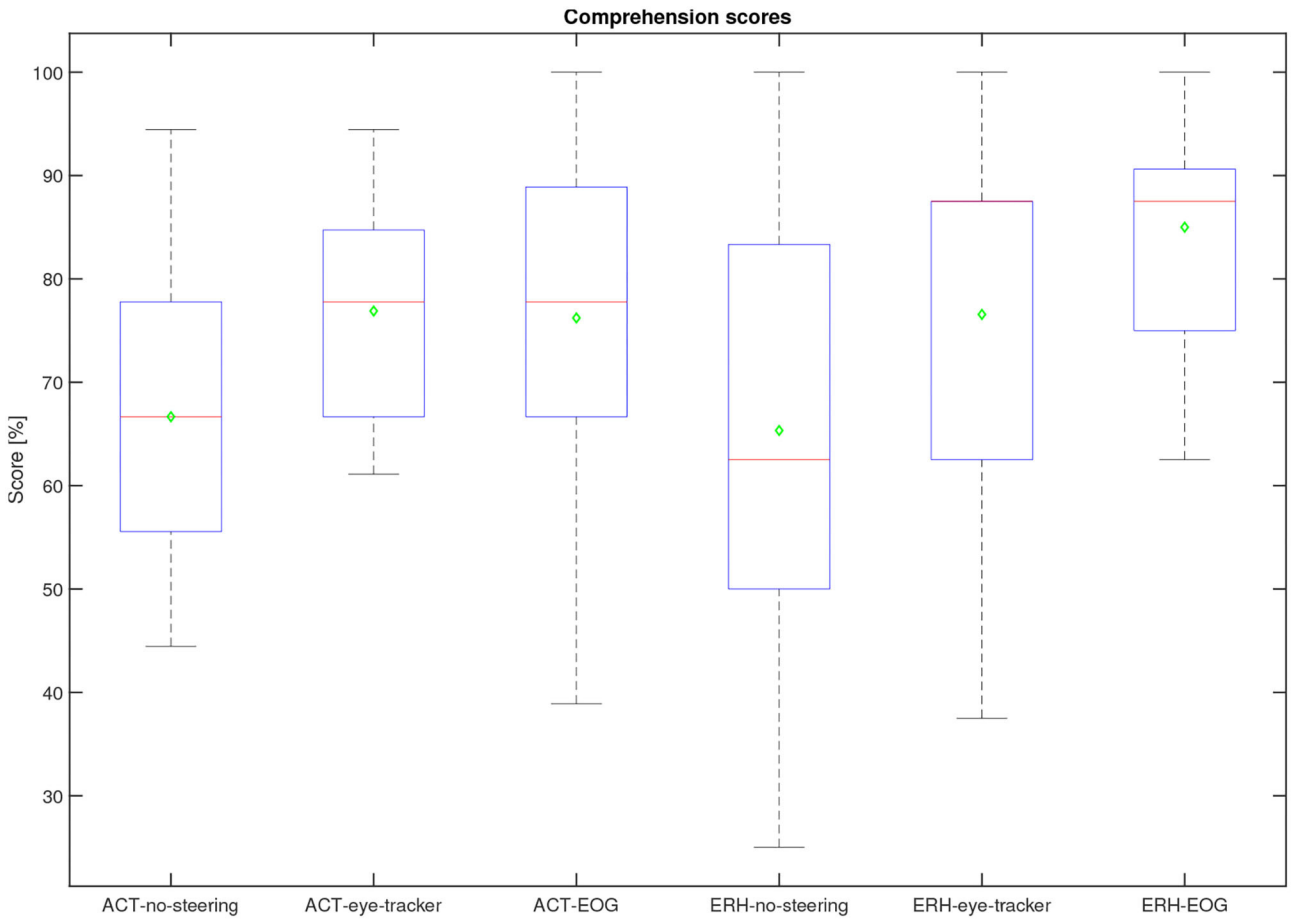


FIGURE 11 | Comprehension performance scores for all participants per condition. The median is the red line, the mean is a green diamond, the box represents the 25 and 75*th* percentile, the whiskers extend to the extreme values, and the red crosses represent individual outliers.

eye-tracker condition would perform better than the EOG condition because the EOG signals were not expected to be reasonably free of distortion for all participants, but the eye-tracker signals were. Apart from the possibility that the block of trials used for the in-ear EOG condition was easier to comprehend than those used for the eye-tracker condition, it is also possible that a certain amount of support was given even when the EOG steering did not react to saccades. This is because one speaker channel was always amplified by 6 dB, whether this was the channel for the speaker talking or not. This means that the participants could potentially pick up some keywords from the active speaker that was beneficial for comprehension, even if they did not directly attend to that speaker.

The non-significant difference in comprehension performance between the EOG and eye-tracker conditions suggest that there can be equal benefit from conventional eye-tracking and in-ear EOG steering. Thus, it should not be concluded that a lower attended speaker accuracy in the EOG condition is indicative of lesser comprehension improvement than in the eye-tracker condition, although differences in comprehension difficulties across the blocks used in the two conditions may have counteracted that. Nevertheless, the accuracy of estimating the attended speaker using in-ear EOG could never exceed that of the eye-tracker which is considered the ground truth reference. Due to the lack of a reference for the accuracy of eye-tracking, it cannot be deduced that this type of steering alone improves the comprehension score. However, as noted in Section 3.1, the comprehension score was improved in the EOG condition when the estimated attended speaker was more accurate. This can be taken as indirect support that when the accuracy of the EOG method is more similar to eye-tracker accuracy, then the speech comprehension is improved.

For future research, there are a few directions of particular interest. To fine-tune the experimental setup and to develop new stimuli that represent a variety of real-life communication situations, it is desirable to obtain a better general understanding of the validity of the comprehension paradigm introduced here; trial clip equivalence; and the significance of including visual components. Furthermore, developing a setup that allows us to better understand the interplay between steering and comprehension is desirable. It is likely that a simpler paradigm with more control of

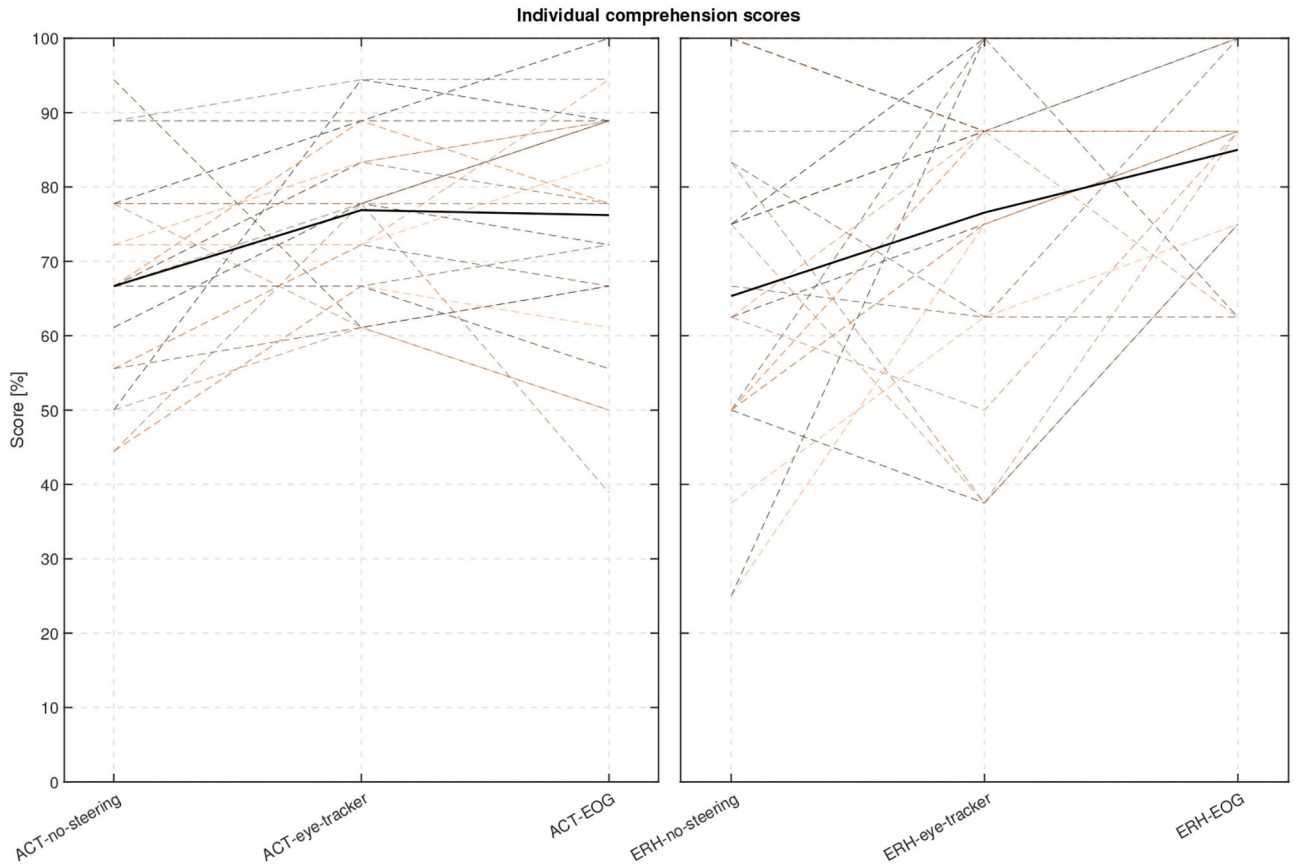


FIGURE 12 | Individual comprehension performance scores for each participant and conditions with the mean scores shown as thick lines.

the task, for instance, speech intelligibility could be an option. These insights could guide the design for more efficient and intuitive attention switching algorithms and test paradigms.

5. CONCLUSION

A method for visual attention estimation using in-ear EOG was evaluated on hearing-impaired participants using an audio-visual dialog presented in noise. Particularly, a novel calibration procedure used to identify the strongest EOG signal available for estimating the attended speaker was investigated for accuracy against that obtained with a conventional eye-tracker. Comprehension performance with the two methods was also measured. The causal relationships found between the strength of various calibration metrics and greater attention estimation accuracy and better speech comprehension are highly encouraging and show great potential for utilizing in-ear EOG in hearing devices to steer signal processing strategies targeting the signal of interest.

DATA AVAILABILITY STATEMENT

The datasets generated and analyzed for this study can be obtained upon request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the Capital Region in Denmark. 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

The majority of manuscript preparation was done by MAS with assistance from all authors. MAS, SR-G, and MA developed the technical setup, software, and analyzed the data. SR-G and MA carried out the experiments and

data collection. All authors contributed to the stimulus material development.

FUNDING

This work was financially supported by the Swedish Research Council (Vetenskapsrådet, VR 2017-06092 Mekanismer och behandling vid åldersrelaterad hörselnedsättning).

ACKNOWLEDGMENTS

The authors are grateful for the kind support from Allan Thiell Sørensen who coordinated the project, Alejandro Lopez Valdez who suggested several improvements to the EOG setup, and Thomas Lunner who made valuable input to the methodology of preceding experiments.

REFERENCES

Baker, R., and Hazan, V. (2011). DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behav. Res. Methods* 43, 761–770. doi: 10.3758/s13428-011-0075-y

Belkhiria, C., and Peysakhovich, V. (2021). EOG metrics for cognitive workload detection. *Procedia Comput. Sci.* 192, 1875–1884. doi: 10.1016/j.procs.2021.08.193

Best, V., Roverud, E., Streeter, T., Mason, C. R., and Gerald Kidd, J. (2017). The benefit of a visually guided beamformer in a dynamic speech task. *Trends Hear.* 21, 2331216517722304. doi: 10.1177/2331216517722304

Bo Nielsen, J., Dau, T., and Neher, T. (2014). A danish open-set speech corpus for competing-speech studies. *J. Acoust. Soc. Am.* 135, 407–420. doi: 10.1121/1.4835935

Braun, S., Zhou, W., and Habets, E. A. P. (2015). "Narrowband direction-of-arrival estimation for binaural hearing aids using relative transfer functions," in *2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)* (New Paltz, NY: IEEE), 1–5.

Cabella, T. (2021). *Audiovisual listening in hearing-impaired adults* (Master's thesis). University of Glasgow.

Carlile, S., and Keidser, G. (2020). Conversational interaction is the brain in action: implications for the evaluation of hearing and hearing interventions. *Ear Hear.* 41, 56s-67s. doi: 10.1097/AUD.0000000000000939

Chen, M., Anzai, D., Wang, J., Terado, T., and Fischer, G. (2019). Two improving methods of EOG-based eye movement detection for hci. *IEEE Trans. Electron. Inform. Syst.* 139, 1474–1480. doi: 10.1541/ieejeiss.139.1474

Favre-Félix, A., Graversen, C., Bhuiyan, T. A., Skoglund, M. A., Rotger-Griful, S., Rank, M. L., et al. (2019). Absolute eye gaze estimation with biosensors in hearing aids. *Front. Neurosci.* 13, 1294. doi: 10.3389/fnins.2019.01294

Favre-Félix, A., Graversen, C., Dau, T., and Lunner, T. (2017). "Real-time estimation of eye gaze by in-ear electrodes," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Jeju: IEEE), 4086–4089.

Favre-Felix, A., Graversen, C., Hietkamp, R. K., Dau, T., and Lunner, T. (2018). Improving speech intelligibility by hearing aid eye-gaze steering: conditions with head fixated in a multitalker environment. *Trends Hear.* 22, 2331216518814388. doi: 10.1177/2331216518814388

Fiedler, L., Obleser, J., Lunner, T., and Graversen, C. (2016). "Ear-EEG allows extraction of neural responses in challenging listening scenarios – A future technology for hearing aids?" in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Orlando, FL: IEEE), 5697–5700.

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

Grimm, G., Kayser, H., Hendrikse, M., and Hohmann, V. (2018). "A gaze-based attention model for spatially-aware hearing aids," in *Speech Communication; 13th ITG-Symposium (VDE)* (Oldenburg, Germany), 1–5.

Gunawardane, P. D. S. H., MacNeil, R. R., Zhao, L., Enns, J. T., de Silva, C. W., and Chiao, M. (2021). A fusion algorithm for saccade eye movement enhancement with EOG and lumped-element models. *IEEE Trans. Biomed. Eng.* 68, 3048–3058. doi: 10.1109/TBME.2021.3062256

Hadley, L. V., Brimijoin, W. O., and Whitmer, W. M. (2019). Speech, movement, and gaze behaviours during dyadic conversation in noise. *Scientific Rep.* 9, 10451. doi: 10.1038/s41598-019-46416-0

Hládek, L., Porr, B., and Brimijoin, W. O. (2018). Real-time estimation of horizontal gaze angle by saccade integration using in-ear electrooculography. *PLoS ONE* 13, e0190420. doi: 10.1371/journal.pone.0190420

Huigen, E., Peper, A., and Grimbergen, C. A. (2002). Investigation into the origin of the noise of surface electrodes. *Med. Biol. Eng. Comput.* 40, 332–338. doi: 10.1007/BF02344216

Kappel, S. L., and Kidmose, P. (2015). "Study of impedance spectra for dry and wet eareeg electrodes," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Milan: IEEE), 3161–3164.

Kappel, S. L., and Kidmose, P. (2018). "Real-life dry-contact ear-EEG," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Honolulu, HI: IEEE), 5470–5474.

Kappel, S. L., Rank, M. L., Toft, H. O., Andersen, M., and Kidmose, P. (2018). Dry-contact electrode ear-EEG. *IEEE Trans. Biomed. Eng.* 66, 150–158. doi: 10.1109/TBME.2018.2835778

Kastrati, A., Plomecka, M. B., Pascual, D., Wolf, L., Gillioz, V., Wattenhofer, R., et al. (2021). "EEGEyeNet: a simultaneous electroencephalography and eye-tracking dataset and benchmark for eye movement prediction," in *Thirty-Fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 1)*. Available online at: https://neurips.cc/Conferences/2021

Keidser, G., Naylor, G., Brungart, D. S., Caduff, A., Campos, J., Carlile, S., et al. (2020). The quest for ecological validity in hearing science: what it is, why it matters, and how to advance it. *Ear Hear.* 41(Suppl. 1), 5s-19s. doi: 10.1097/AUD.0000000000000944

Llorach, G., Grimm, G., Hendrikse, M. M., and Hohmann, V. (2018). "Towards realistic immersive audiovisual simulations for hearing research: capture, virtual scenes and reproduction," in *Proceedings of the 2018 Workshop on Audio-Visual Scene Understanding for Immersive Multimedia* (Seoul, Republic of Korea), 33–40.

Lunner, T., Alickovic, E., Graversen, C., Ng, E. H. N., Wendt, D., and Keidser, G. (2020). Three new outcome measures that tap into cognitive processes required for real-life communication. *Ear Hear.* 41(Suppl. 1), 39s-47s. doi: 10.1097/AUD.0000000000000941

Manabe, H., and Fukumoto, M. (2006). "Full-time wearable headphone-type gaze detector," in *CHI '06 Extended Abstracts on Human Factors in Computing Systems, CHI EA '06* (New York, NY: Association for Computing Machinery), 1073–1078.

Manabe, H., Fukumoto, M., and Yagi, T. (2013). "Automatic drift calibration for eog-based gaze input interface," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Osaka: IEEE), 53–56.

Mehra, R., Brimijoin, O., Robinson, P., and Lunner, T. (2020). Potential of augmented reality platforms to improve individual hearing aids and to support more ecologically valid research. *Ear Hear.* 41(Suppl. 1), 140s-146s. doi: 10.1097/AUD.0000000000000961

Moore, B., and Glasberg, B. (1998). Use of a loudness model for hearing-aid fitting. I. Linear hearing aids. *Br. J. Audiol.* 32, 317–335. doi: 10.3109/03005364000000083

O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., et al. (2014). Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cereb. Cortex* 25, 1697–1706. doi: 10.1093/cercor/bht355

Pomper, U., and Chait, M. (2017). The impact of visual gaze direction on auditory object tracking. *Sci. Rep.* 7, 1–16. doi: 10.1038/s41598-017-04475-1

Roverud, E., Best, V., Mason, C., Streeter, T., and Kidd, G. (2017). Evaluating the performance of a visually guided hearing aid using a dynamic auditory-

visual word congruence task. *Ear. Hear.* 39, 1. doi: 10.1097/AUD.0000000000000532

Slaney, M., Lyon, R. F., Garcia, R., Kemler, B., Gnegy, C., Wilson, K., et al. (2020). Auditory measures for the next billion users. *Ear Hear.* 41(Suppl. 1), 131s-139s. doi: 10.1097/AUD.0000000000000955

Zohourian, M., Enzner, G., and Martin, R. (2018). Binaural speaker localization integrated into an adaptive beamformer for hearing aids. *IEEE/ACM Trans. Audio Speech Lang. Process.* 26, 515–528. doi: 10.1109/TASLP.2017.2782491

Conflict of Interest: MAS, MMS, GK, and SR-G were employed by Oticon A/S, and MA and MR were employed by T&W Engineering A/S.

The reviewer VH declared a past co-authorship with one of the authors GK to the handling Editor.

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

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 25 August 2022
DOI 10.3389/fnins.2022.973735

OPEN ACCESS

EDITED BY
Preben Kidmose,
Aarhus University, Denmark

REVIEWED BY
Betul Cicek Cinar,
Hacettepe University, Turkey
Renjie Chai,
Southeast University, China
Ye Tao,
First Affiliated Hospital of Anhui Medical University, China

*CORRESPONDENCE
Shouqin Zhao
shouqinzhao01@163.com
Shusheng Gong
gongss@ccmu.edu.cn

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

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

RECEIVED 20 June 2022
ACCEPTED 05 August 2022
PUBLISHED 25 August 2022

CITATION
Liu Y, Zhao C, Yang L, Chen P, Yang J, Wang D, Ren R, Li Y, Zhao S and Gong S (2022) Characteristics of sound localization in children with unilateral microtia and atresia and predictors of localization improvement when using a bone conduction device. *Front. Neurosci.* 16:973735. doi: 10.3389/fnins.2022.973735

Characteristics of sound localization in children with unilateral microtia and atresia and predictors of localization improvement when using a bone conduction device

Yujie Liu[1†], Chunli Zhao[2†], Lin Yang[1], Peiwei Chen[1], Jinsong Yang[1], Danni Wang[1], Ran Ren[1], Ying Li[1], Shouqin Zhao[1*] and Shusheng Gong[2*]

[1]Ministry of Education Key Laboratory of Otolaryngology Head and Neck Surgery, Department of Otolaryngology Head and Neck Surgery, Beijing Tongren Hospital, Capital Medical University, Beijing, China, [2]Department of Otolaryngology Head and Neck Surgery, Beijing Friendship Hospital, Capital Medical University, Beijing, China

This study aimed to determine the characteristics of sound localization in children with unilateral microtia and atresia (UMA) and the influence of a non-surgical bone conduction device (BCD). Hearing benefits were evaluated by the word recognition score (WRS), speech reception threshold, the international outcome inventory for hearing aids (IOI-HA), and the Speech, Spatial, and Qualities of Hearing Test for Parent (SSQ-P). Sound localization was measured using broadband noise stimuli randomly played from seven loudspeakers at different stimulus levels [65, 70, and 75 dB sound pressure levels (SPLs)]. The average unaided WRS and speech-to-noise ratio (SNR) for UMA patients was 18.27 ± 14.63 % and −5 ± 1.18 dB SPL, and the average aided WRS and SNR conspicuously changed to 85.45 ± 7.38 % and −7.73 ± 1.42 dB SPL, respectively. The mean IOI-HA score was 4.57 ± 0.73. Compared to the unaided condition, the mean SSQ-P score in each domain improved from 7.08 ± 2.5, 4.86 ± 2.27, and 6.59 ± 1.4 to 8.72 ± 0.95, 7.61 ± 1.52, and 8.55 ± 1.09, respectively. In the sound localization test, some children with UMA were able to detect sound sources quite well and the sound localization abilities did not deteriorate with the non-surgical BCD. Our study concludes that for children with UMA, the non-surgical BCD provided a definite benefit on speech recognition and high satisfaction without deteriorating their sound localization abilities. It is an efficient and safe solution for the early hearing intervention of these patients.

KEYWORDS

unilateral, microtia and atresia, congenital conductive hearing loss, speech perception, sound localization, bone conduction device

Introduction

Hearing loss is a major global problem and was described as an epidemic of the twenty-first century by the World Health Organization [WHO] (2021). Hearing loss can be congenital or acquired, with possible etiologies including genetic causes (Hong et al., 2022; Tao et al., 2022), infections (Shahar-Nissan et al., 2022), excessive noise (Jiang et al., 2021), ototoxic drugs (Fu et al., 2022; Wang et al., 2022; Zhang et al., 2022), and aging (He et al., 2021). For patients with profound bilateral hearing impairment, treatment is always offered early in life, but for patients with unilateral form, the hearing intervention tends to be delayed, as the normal hearing (NH) ear provides enough hearing cues for basic speech understanding. However, functional deficits of disability in speech recognition and inaccuracy of sound localization have been reported in patients with unilateral hearing loss, and they may also experience an apparent handicap in academic performance and social interactions (van Wieringen et al., 2019; Okada et al., 2020).

Microtia and atresia, a developmental malformation of the middle and external ear, is a common cause of congenital conductive hearing loss. Two-thirds of patients with microtia and atresia experience the unilateral form [i.e., unilateral microtia and atresia (UMA)] with unilateral conductive hearing loss (UCHL) (Luquetti et al., 2012; Bartel-Friedrich, 2015). Common treatment options for patients with UMA include traditional canaloplasty, active middle ear implants, bone conduction implants, and non-surgical bone conduction devices (BCDs). For children with UMA who are not willing to undergo surgery or who have not reached the age for surgery, non-surgical BCDs represent an important transition intervention (Liu et al., 2017). Non-surgical BCDs can provide evident speech recognition-related benefits to patients with UCHL; however, whether these patients can achieve more accurate sound localization after receiving such interventions remains disputed (Kunst et al., 2008; Yu et al., 2014; Liu et al., 2017; Vyskocil et al., 2017). Thus, doctors and parents often face the dilemma of whether a BCD should be selected for these children at an early age.

Sound localization is a complex process that relies on the computation and integration of multiple spatial cues at the level of the auditory pathway (Tillein et al., 2016; Risoud et al., 2018; Wood et al., 2019). For patients having acquired UCHL with a mature auditory system, definite improvement of sound localization ability was observed after hearing intervention (Agterberg et al., 2011, 2012). Regarding congenital UCHL, the results seem to be contentious. Some studies have reported remarkable improvements in horizontal spatial hearing in patients with congenital UCHL aided with BCD, despite the inherent problems of time delay and cross-hearing (Nelissen et al., 2016; Vyskocil et al., 2017). In contrast, other studies have suggested that congenital UCHL cannot benefit from BCDs in terms of horizontal spatial hearing abilities (Kunst et al., 2008; Weiss et al., 2017). They maintained that listeners with congenital UCHL might have adapted to their hearing impairment as they learned to rely on the spectral shape cues and ambiguous monaural head shadow effect (HSE) cues, which had developed during the long-term unilateral hearing deprivation (Van Wanrooij and Van Opstal, 2004; Vogt et al., 2020). When aided with a BCD, such listening cues might be distorted sharply, thus jeopardizing the original directional hearing. Given the uncertain benefits of hearing amplification and non-aesthetic reasons, studies concerning the sound localization ability of children with UMA are limited by heterogeneous patient populations, varying in study design and audiological test results. How bone conduction (BC) stimulation affects spatial hearing abilities and the predictive factors that may affect the degree of the benefit provided by BCDs are still unknown.

Currently, there is no research investigating the characteristics of sound localization and the effects of non-surgical BCDs in school-aged children with UMA. This study had three primary objectives: to detect the hearing benefits of a BCD on speech perception and subjective satisfaction in children with congenital UMA; to compare characteristics of sound localization in children with congenital UMA and children with NH, as well as acquired UCHL; to investigate whether the use of BCD would be detrimental to the original sound localization of children with UMA and reveal predictive factors for the improvement of sound localization accuracy after using a BCD.

Materials and methods

Ethics

Ethical approval was given by the medical committee of Beijing Tongren Hospital, Capital Medical University (TRECKY2018-067). Written informed consent for participation was obtained from the parents and guardians of the participants.

Abbreviations: NH, normal hearing; UMA, unilateral microtia and atresia; UCHL, unilateral conductive hearing loss; BCD, bone conduction devices; HSE, head shadow effect; BC, bone Conduction; SD, standard deviation; HL, hearing level; SPL, sound pressure level; UP, the unplugged condition; P, the plugged condition; FHG, functional hearing gain; WRS, word recognition score; MSTM, Mandarin Speech Test Materials; SRT, speech reception threshold; SSN, spectrum-shaped noise; SNR, speech-to-noise ratio; IOI-HA, the International Outcome Inventory for Hearing Aids; SSQ-P, the Speech, Spatial, and Qualities of Hearing Test for Parent; MAE, mean absolute error; g/gain, response gain; b, response bias; r^2, R square.

Participants

Eleven children (mean ± SD: 7.45 ± 1.81 years) who had UMA and congenital UCHL were included. All patients had NH in one ear [hearing thresholds, ≤ 25 dB hearing level (HL) across 0.5–4 kHz] and pure conductive hearing loss in the impaired ear (air-bone gap, ≥ 25 dB HL, BC thresholds of ≤ 25 dB HL across 0.25–4 kHz). For comparative purposes, eight boys and three girls aged 6–12 years who had bilateral NH were recruited as control listeners. All control listeners had bilateral air and BC hearing thresholds ≤ 25 dB HL across frequencies of 0.5–4 kHz. Detailed demographic data are summarized in **Table 1**.

Device and listening conditions

The BCD used in the current study was a non-surgical adhesive device (ADHEAR; MED-EL, Innsbruck, Austria). All devices were set in the omnidirectional mode for all experimental conditions, and the volume was adjusted based on patients' preferences. The system fittings of the ADHEAR did not change during all experiments.

Children with UMA were tested with the BCD off (unaided condition) and on (aided condition) (**Figure 1A**). Children with NH were measured with both ears unplugged (UP condition) as normal controls. When measuring sound localization, control listeners were also tested with plugging (P condition) to stimulate an acquired UCHL to reveal the difference in directional hearing between the acquired UCHL and UMA (congenital UCHL). Plugging was performed by covering an ear with an earmuff (Peltor X5A; 3M Company, MN, United States), along with a foam earplug (E-A-R soft; 3M Company, MN, United States) inserted into the external auditory canal. The plugging provided a mean attenuation of 40.22 dB ± 2.29 dB HL, from 0.5 to 4 kHz (measured by audiometric threshold shifts) in the sound field.

For the UMA group, unaided audiological tests were performed on the day they received the BCD and aided audiological tests were measured after a mean period of 9.27 ± 1.85 weeks. For the NH group, all tests were performed in one visit.

Setup and stimuli

All tests were conducted in a double-walled soundproof laboratory. Participants sat in a chair placed 1 m in front of seven loudspeakers. Sound field hearing thresholds were obtained by warble tones for octave frequencies across 0.25–4 kHz in dB HL. Speech perception under quiet was measured by the word recognition score [WRS (%)] of the Mandarin speech test materials (MSTMs) (Wang et al., 2007) at 65 dB sound pressure level (SPL). Speech perception in noise was measured by the speech reception threshold (SRT) of the MSTMs. The spectrum-shaped noise (SSN) was set at 65 dB SPL, and the speech signal started at 0 dB speech-to-noise ratio (SNR), with the following disyllables changing adaptively in 2 dB SPL steps as the participants responded. The SRT was defined as the speech signal level presented when a participant identified 50% of the words correctly. The SNR was calculated as the difference between the SRT and SSN.

Sound localization was measured in a double-walled soundproof laboratory with seven audiometric loudspeakers placed at 30° intervals in a semicircle within the horizontal plane (± 90°, azimuth) (**Figure 1B**). Broadband noise (0.5–20 kHz), with a duration of 1 s, was randomly played at three different sound levels (65, 70, and 75 dB SPL). During the test, each loudspeaker was randomly presented twice at each sound level burst; thus, 42 stimuli were included in each test. The participants sat comfortably in a chair located 1 m in front of the loudspeaker, facing and fixating the loudspeaker at 0°, azimuth. They were not permitted to move their heads when the noise bursts were presented. After the loudspeaker finished each presentation, participants were allowed to indicate the orientation and could turn their heads to look at the number of the loudspeakers that they considered to be the source of the burst.

To familiarize the participants with the experiment before the formal sound localization tests, a brief block of 12 broadband stimuli was presented. They were instructed to localize the stimuli as fast as possible, and no feedback was provided throughout the training to avoid the influence of learning in the formal test.

Subject satisfaction

The subjective satisfaction was measured with two questionnaires, the Chinese version of the International Outcome Inventory for Hearing Aids (IOI-HA) (Liu et al., 2011) and the Speech, Spatial, and Qualities of Hearing Test for Parent (SSQ-P) (Gao et al., 2022), which were handed out to patients' parents at the end of the follow-up. The IOI-HA consists of seven items: daily use, benefit, residual activity limitations, satisfaction, residual participation restrictions, impact on others, and quality of life. Each answer is rated on a scale from 0 to 5, with higher ratings reflecting better outcomes (or fewer residual difficulties). The SSQ-P across three domains: speech, spatial hearing, and qualities of hearing, with higher scores in each subdomain representing higher satisfaction. To evaluate the hearing impairment of patients with UMA in unaided conditions, the SSQ-P was also handed out to their parents before they were equipped with the ADHEAR.

TABLE 1 The demographic data of 11 patients with UMA and 11 children with NH.

Participant number	Sex	Age (years)	Side of impaired/plugged	Etiology	Follow-up time (weeks)
P1	M	9	R	Atresia	8
P2	F	7	R	Atresia	13
P3	M	8	L	Stenosis	9
P4	M	8	R	Atresia	8
P5	F	5	L	Atresia	8
P6	M	7	R	Atresia	12
P7	F	11	L	Stenosis	8
P8	M	6	R	Atresia	9
P9	M	7	L	Atresia	8
P10	M	5	R	Stenosis	11
P11	M	9	L	Atresia	8
Mean ± *SD*	–	7.45 ± 1.81			9.27 ± 1.85
N1	M	8	R	–	–
N2	M	9	L	–	–
N3	F	11	R	–	–
N4	M	11	R	–	–
N5	F	7	L	–	–
N6	M	9	R	–	–
N7	F	6	L	–	–
N8	M	7	R	–	–
N9	M	8	L	–	–
N10	M	12	L	–	–
N11	M	10	R	–	–
Mean ± *SD*	–	8.91 ± 1.92	–	–	–

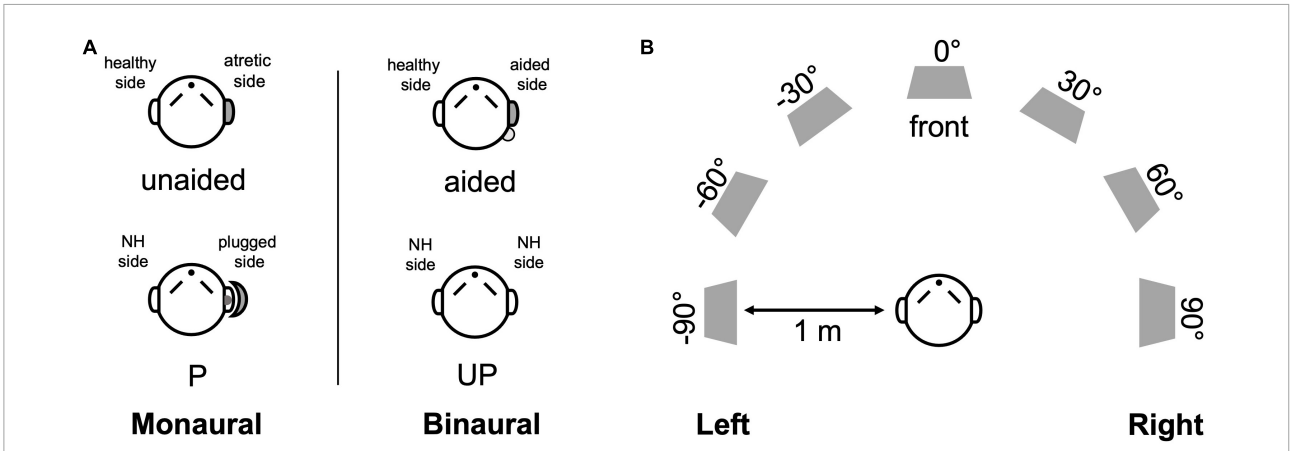


FIGURE 1
Test setup and listening conditions. **(A)** The monaural (left) and binaural listening (right) conditions are designed for UMA patients (unaided and aided conditions) and controls (P and UP conditions). **(B)** Seven loudspeakers were placed at 30° intervals in a semicircle in a double-walled, soundproof laboratory. NH, normal hearing; P, the plugged condition; UP, the unplugged condition.

Data analysis

$$MAE = \frac{\sum_{i=1}^{n} \left| \alpha_{RESP}^{i} - \alpha_{TARG}^{i} \right|}{n} \quad (1)$$

$$\alpha_{RESP} = gain \cdot \alpha_{TARG} + b \quad (2)$$

The mean absolute error (MAE) was calculated using Equation 1 to assess the sound localization accuracy under different conditions, where the α_{RESP} and α_{TARG} referred

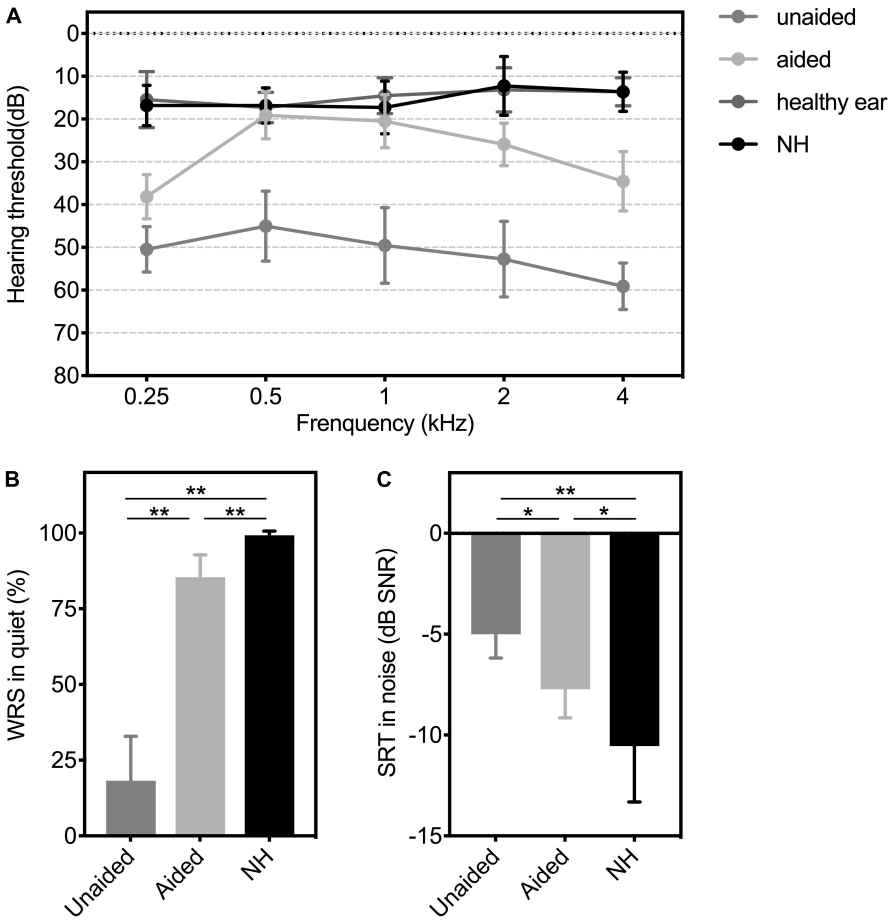


FIGURE 2
(A) Hearing thresholds, **(B)** WRS in quiet, and **(C)** SRT in the noise of patients with UMA in unaided and aided conditions, as well as the controls. Group means are presented as mean ± *SD*. Significant differences, *$p < 0.05$, **$p < 0.01$. WRS, word recognition score; SRT, speech reception threshold; SNR, speech-to-noise ratio; *SD*, standard deviation; Unaided, the unaided condition of UMA; Aided, the aided condition of UMA; healthy ear: the healthy ear of patients with UMA; NH, normal hearing group.

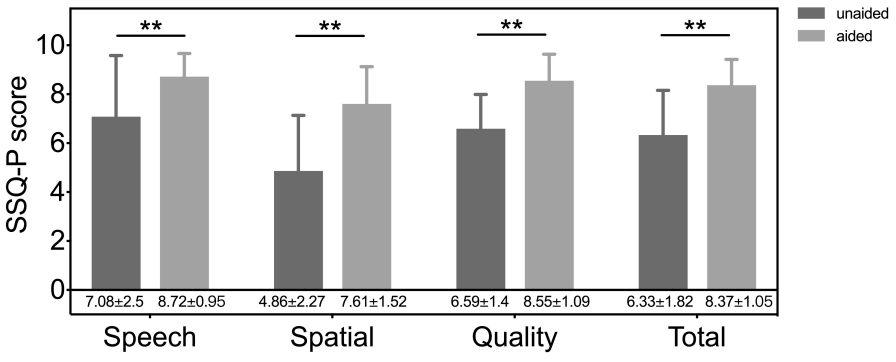


FIGURE 3
The results of speech, spatial and qualities of hearing scale for parents (SSQ-p) without (unaided) and with (aided) a BCD. Significant differences, **$p < 0.01$. Unaided, the unaided condition of UMA; aided, the aided condition of UMA.

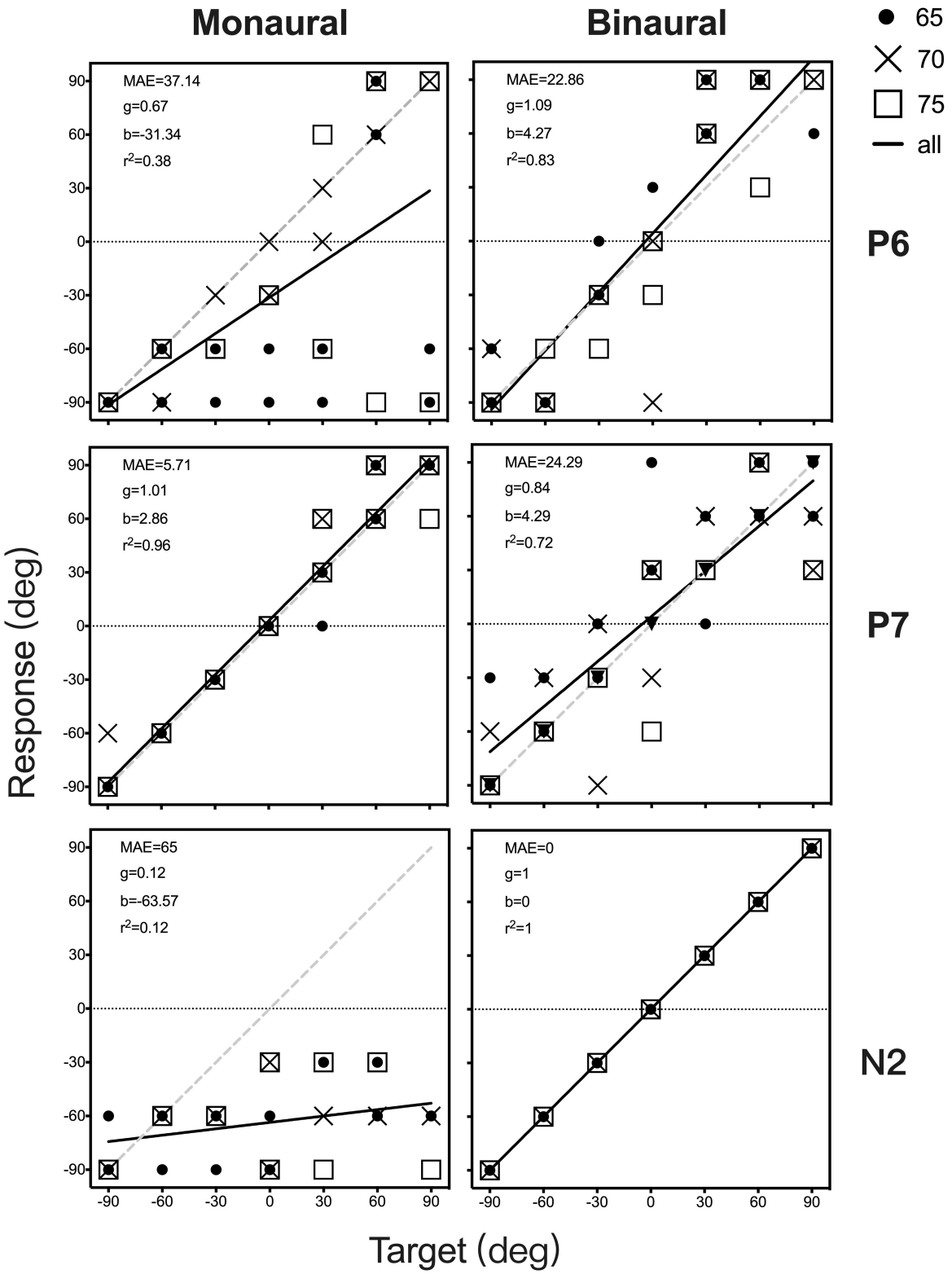


FIGURE 4
Sound localization target-response plots of two patients (P6 and P7) and one control (N2) in monaural (unaided and P, left) and binaural (aided and UP, right) listening conditions. Stimulus sound levels are indicated by black circle (65 dB SPL), and cross data (70 dB SPL), and white square points (75 dB SPL). Best-fit linear regression is indicated by a black line. For participants with an ideal optimal localization ability, gain is 1, whereas MAE and *b* are 0. MAE, mean absolute error; g, response gain; b, response bias; r^2, R square.

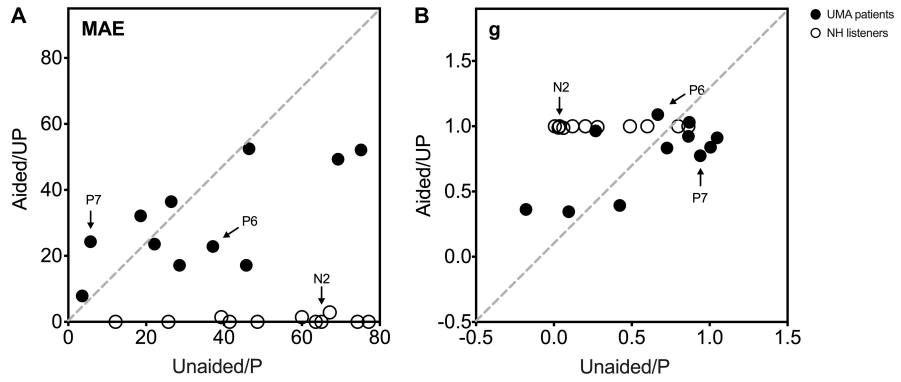


FIGURE 5
(A) The MAE and **(B)** g under monaural listening conditions (unaided and P, Y-axis) are plotted against those under binaural listening conditions (aided and UP, x-axis). Black circle data points indicate 11 UMA children, and white circle data points indicate the controls. The two UMA children and one control depicted in **Figure 4** are marked in this figure (P6, P7, and N2). An MAE near 0 and a gain near 1 demonstrate a close-to-normal sound localization performance. Data of participants with the same sound localization performance when listening monaurally and binaurally are displayed on the gray dotted diagonal. A data point below the diagonal in **(A)** and above the diagonal in **(B)** represents a better sound localization performance when listening under binaural conditions than under monaural conditions. UMA, unilateral microtia and atresia; NH, normal hearing group; MAE, mean absolute error; g, response gain; Unaided, the unaided condition of UMA; Aided, the aided condition of UMA; P, the plugged condition of controls; UP, the unplugged condition of controls.

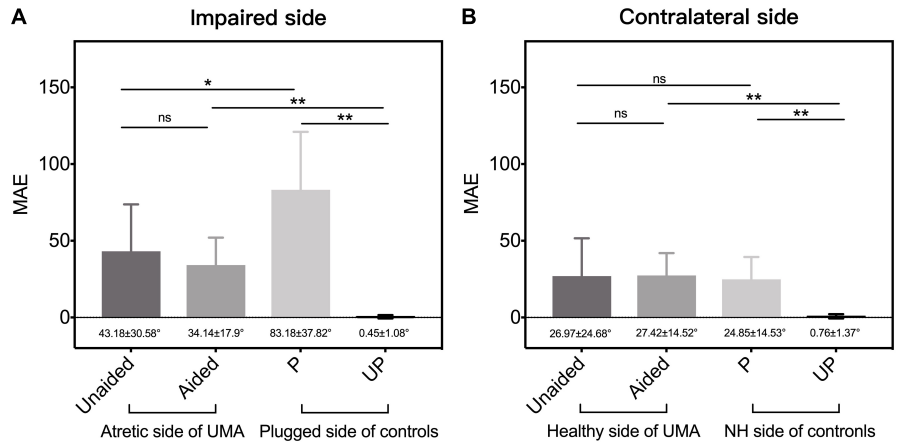


FIGURE 6
The mean MAE of patients with UMA and the controls, respectively, on **(A)** the impaired (including the atretic side of UMA and the plugged side of controls) side and the **(B)** contralateral (the healthy side of UMA and the unplugged side of controls) side. Error bars represent mean ± *SD*. Significant differences, *$p < 0.05$, **$p < 0.01$. ns, not significant; UMA, unilateral microtia and atresia; MAE, mean absolute error; Unaided, the unaided condition of UMA; Aided, the aided condition of UMA; P, the plugged condition of controls; UP, the unplugged condition of controls; NH, normal hearing; *SD*, standard deviation.

to the response azimuth and target azimuth (both in degrees). Additionally, the best linear fit of the target-response relationship for each participant was also computed using Equation 2, where g is the response gain (slope, dimensionless), and b is the response bias (offset in degrees). In this study, the right side was defined as the impaired side; therefore, azimuth coordinates for patients with left ear impairment and controls with NH with left ears plugged were inverted.

Paired and independent t-tests were conducted to evaluate differences under different test conditions. The Mann-Whitney U-test was used to compare the difference in the unaided, aided, and delta MAEs (delta MAE = aided MAE - unaided MAE) between the groups of different sexes, sides of impairment, and etiologies. Spearman correlation analysis was conducted to analyze the correlations between continuous variables (age and follow-up time) and the unaided, aided, and delta MAEs, respectively. The p-values of < 0.05 and < 0.01 were

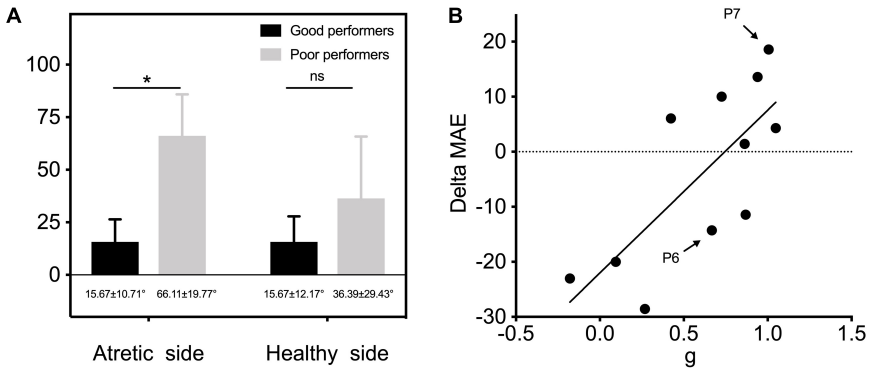


FIGURE 7
(A) MAE outcomes of two subgroups of good performers (n = 5; gain > 0.75) and poor performers (n = 6; gain ≤ 0.75) were separately compared on the atretic and healthy sides. **(B)** Individual data of gain of UMA children in unaided conditions are plotted as a function of delta MAE between unaided and aided conditions. The linear regression was conducted to explore the predictive effect of gain on the benefits of sound localization accuracy by fitting BCDs. P6 and P7 depicted in **Figure 4** are marked in this figure. Delta MAE, aided MAE—unaided MAE. Significant differences, *p < 0.05. ns, not significant; MAE, mean absolute error; g, response gain.

considered statistically significant. SPSS version 26.0 (IBM Corp., Armonk, NY, United States) and GraphPad Prism version 8.0 (GraphPad, San Diego, CA) were used to analyze the data and draw diagrams.

Results

Hearing benefits

For patients with UMA, the mean hearing threshold of the healthy ear was 14.82 ± 3.82 dB HL. The unaided hearing threshold was 51.36 ± 5.02 dB HL and significantly improved to 27.64 ± 2.38 dB HL with a mean functional hearing gain (FHG) of 23.73 ± 3.47 dB HL (p < 0.01). For the NH comparison group, the mean hearing threshold was 15.36 ± 3.88 dB HL. The aided mean hearing threshold of the UMA group was still higher (worse) than the mean hearing threshold of the NH group (p < 0.01). Detailed outcomes across 0.5–4 kHz are depicted in **Figure 2A**.

The average unaided WRS and SNR for patients with UMA were 18.27 ± 14.63 % and −5 ± 1.18 dB SPL, respectively, whereas the average aided WRS and SNR conspicuously changed to 85.45 ± 7.38 % and −7.73 ± 1.42 dB SPL, respectively (WRS: p < 0.01; SNR: p < 0.05). The mean WRS and SNR of the comparison group were 99.27 ± 1.35 % and −10.55 ± 2.77 dB SPL, respectively. **Figures 2B,C** show significant differences in the speech levels between patients aided with BCDs and their peers with NH (WRS: p < 0.01; SNR: p < 0.05).

The mean overall IOI-HA score was 4.57 ± 0.73. A score > 3 per item, defined as a benefit from the BCD, was found for nearly all participants. The mean score for items 1–7 were 3.86 ± 0.31, 3.57 ± 0.25, 4.29 ± 0.19, 3.79 ± 0.24, 4.14 ± 0.14, 4.64 ± 0.13, and 4 ± 0.23, respectively. The results of the SSQ-P without (unaided) and with (aided) BCD are presented in **Figure 3**, and significant improvements of subjective satisfaction with the ADHEAR were found in each subdomain and the total rating (all p < 0.01).

Sound localization in patients with unilateral microtia and atresia and stimulated unilateral conductive hearing loss

Figure 4 shows the individual sound localization target-response plots for two children with UMA (P6 and P7) and one control (N2) under monaural (unaided and P, left column) and binaural (aided and UP, right column) listening conditions. Under the unaided condition, P6 showed a poor localization ability and perceived most stimuli from the healthy ear side. However, P7 exhibited relatively better sound localization accuracy than P6. When aided with the BCD, the sound localization accuracy improved in P6 (delta gain = 0.422, delta MAE = -14.28°), and the application of the BCD led to a decrease in sound localization accuracy in P7 (delta gain = -0.167, delta MAE = 18.58°). Under the P condition, all data points of N2 fell along the diagonal dotted line, indicating a sharply deteriorated localization performance with the data points spread larger on the NH side (gain = 0.12, MAE = 65°).

Individual data on sex, age, MAE, gain, bias, and r^2 for all participants are presented in **Supplementary Table 1**. The MAE and gain under monaural listening conditions (unaided and P) are plotted against those under binaural listening conditions (aided and UP) in **Figure 5**. Varying

sound localization performance was observed in the 11 children with UMA under the unaided condition. When the mean gain and MAE of all children with UMA were compared between the unaided and aided conditions, no significant differences were identified (gain: $p = 0.104$, MAE: $p = 0.436$). Control listeners showed good sound localization performance in the UP condition. All exhibited considerable deterioration after being plugged (gain: $p < 0.01$; MAE: $p < 0.01$), with most of them being unable to localize the stimuli presented from the plugged side. Although no significant difference was observed between the unaided and P conditions (gain: $p = 0.073$; MAE: $p = 0.073$), the results indicated that children with UMA showed better sound localization performance (smaller MAE) than the stimulated acquired UCHL listeners. This phenomenon might be related to the adaptation to congenital unilateral asymmetric hearing loss; however, the benefit of this adaptation was insufficient for children with UMA to localize sound as accurately as the normal controls did.

Influence of a bone conduction device on sound localization accuracy

To further explore the influence of BCDs on the localization ability of patients with UMA. The localization accuracy of the patients with UMA and the controls were calculated on the impaired (including the atretic and plugged) and the contralateral (including the healthy and unplugged) side, respectively.

A better sound localization accuracy was observed in children with UMA (43.18 ± 30.58° vs. 83.18 ± 37.82°, $p < 0.05$) on the impaired side than in controls in the P condition (**Figure 6A**). The relative better sound localization accuracy observed in unaided children with UMA, as compared with plugged control listeners, may be attributed to the utilization of distorted remaining binaural cues. For patients aided with a BCD, there was no difference in the MAE between the unaided and aided conditions on the impaired side (43.18 ± 30.58° vs. 34.14 ± 17.9°, $p = 0.303$) or the contralateral side (26.97 ± 24.68° vs. 27.42 ± 14.52°, $p = 0.79$), indicating that the BCD use was not detrimental to the original sound localization ability of the patients with UMA (**Figure 6**).

Prediction of the benefits of sound localization accuracy among bone conduction device users

Our results showed that some patients with congenital UCHL have relatively good monaural directional hearing without any hearing intervention (e.g., P7). All 11 children with UMA were divided into two subgroups of good performers ($n = 5$; gain > 0.75) and poor performers ($n = 6$; gain ≤ 0.75) according to the criterion in Agterberg et al.'s (2019) research. When the MAE outcomes were separately compared bilaterally, a significantly better sound localization accuracy was observed in good performers on the atretic side (15.67° ± 10.71° vs. 66.11° ± 19.77°, $p < 0.05$, **Figure 7A**).

Correlational analysis was further conducted to explore the predictive effect of gain on the benefits of sound localization accuracy by fitting BCDs (delta MAE = aided MAE—unaided MAE, with a smaller delta MAE representing a better improvement in sound localization accuracy). The results revealed an evident relationship between gain and delta MAE ($r^2 = 0.553$, $p < 0.05$), indicating that children with UMA who have poor sound localization performance (lower gain) showed more improvement in sound localization after being fitted with BCDs (**Figure 7B**).

Besides, the Mann-Whitney *U*-test and Spearman correlation analysis regressions were conducted to investigate the influence of sexes, sides of impairment, etiologies, age, and follow-up time on unaided, aided, and delta MAEs in patients with UMA. The results indicated the absence of the main effect (**Supplementary Tables 2**, **3**).

Discussion

Hearing benefits of a bone conduction device

In the present study, the ADHEAR system remarkably improved the hearing thresholds and speech perception under quiet and noisy conditions in patients with UMA. Patients with UMA had a mean FHG of 23.73 ± 3.47 dB HL over a frequency range of 0.5–4 kHz; this result lies in the middle of the range of the previously published data of children wearing the ADHEAR system (17–35.6 dB HL) (Dahm et al., 2019; Neumann et al., 2019). Better speech perception abilities were also achieved in quiet and noisy conditions with high participant satisfaction post-BCD-use.

Sound localization performance of children with unilateral microtia and atresia

Consistent with the findings of previous studies (Agterberg et al., 2019), our results showed an inter-subject variability of directional hearing in children with UMA in unaided conditions. Amongst children with congenital UCHL, good performers might have learned to use remaining binaural difference cues to localize sound sources without hearing amplification, especially when stimuli are presented at an

intensity higher than the audibility of the affected ear (Thompson et al., 2020). Another hypothesis for good directional hearing in patients with UCHL is that some of them may rely on monaural cues to achieve good sound localization performance. Vogt et al. (2020) confirmed that patients with congenital UCHL rely on monaural spectral cues to detect high-frequency sound sources by comparing localization accuracy with and without covering the normal hearing ear pinna. Van Wanrooij and Van Opstal (2004) evaluated nine listeners with chronic unilateral hearing loss through a group of broadband sound stimuli fixed at 60 dB, and the results indicated a strong reliance on the ambiguous HSE in familiar acoustic environments. However, no relationship was found between patients' characteristics and their unaided sound localization performance.

Influence of a bone conduction device on sound localization accuracy

In our study, no significant improvement in sound localization accuracy was observed in children with UMA aided with BCDs. Similar results have also been obtained in previous studies (Kunst et al., 2008; Weiss et al., 2017) regarding the application of bone-anchored hearing aid and Bonebridge (MED-EL, Innsbruck, Austria) in congenital UCHL. The inability to perform binaural hearing may be a consequence of two factors: first, the hearing asymmetry still exists, as the BCD was not able to produce sufficient intensity input to provide the same hearing threshold as that of the ear with NH; second, the processing time delay and inconsistent stimulation are inherent characteristics of BCD signals, and the BC signals with less reliable and constant cues may also prevent children with congenital UCHL from having restored binaural hearing (Sabin et al., 2005). However, other studies reported improvement in sound localization accuracy when a BCD was used in patients with congenital UCHL (Nelissen et al., 2016; Vyskocil et al., 2017; Vogt et al., 2018). In a recent study involving nine children and adolescents with congenital UMA, better spatial hearing accuracy was found when listening through the Bonebridge, suggesting that this benefit is not based on the processing of binaural cues because the improvement was only observed on the impaired ear side (Vogt et al., 2018). Potential explanations for these conflicting observations may be the age gap of the enrolled patients and methodological differences in the procedure. In summary, it is favorable that sound localization abilities of the intact ear did not deteriorate with the cross-hearing of the BCD use, and this result might be a consequence of the insufficient high-frequency sound transmission of BCDs (Dobrev et al., 2019) and does not interfere with the spectral cues from the contralateral healthy ear.

Predictive factors for the benefits of sound localization accuracy among bone conduction device users

As shown in **Figure 7B**, patients with poor unaided spatial hearing (e.g., P6) exhibited more evident improvement (smaller delta MAE) in sound localization accuracy when aided with a BCD. Hence, the original horizontal sound localization performance of listeners with UCHL was a good predictor of their sound localization accuracy under BCD-aided conditions; thus, there is a greater need for early hearing intervention in poor performers who cannot make good use of remaining binaural differences to localize sound sources. As asymmetry hearing induces auditory system reorganization, and animal models of UCHL have shown the structural and functional weakness of the auditory system, thereby affecting binaural hearing integration (Tillein et al., 2016), there seems to be a consensus that early rehabilitation of binaural hearing seems to be better than later rehabilitation (Shirane et al., 2020).

One main limitation of the study is that the factors influencing sound localization accuracy amongst listeners with congenital UCHL are not entirely clear. The small age span in the present study (we mainly included children aged 5–11 years) may be attributed to the absence of a significant correlation between patient characteristics and individual sound localization differentiation. Thus, more factors influencing sound localization performance conflict and the optimal age of BCD use need to be investigated in further studies that include more participants of different ages.

In conclusion, some children with UMA were able to compensate using the remaining distorted binaural cues to detect sound sources, unlike the children with stimulated acquired UCHL; however, this compensating ability was still far worse than children with NH and varied across individuals. As the application of BCD provided a definite benefit on speech recognition abilities and high participant satisfaction, it is recommended that children, particularly those with poor sound localization performance, should be fitted with non-surgical BCDs at an early age.

Data availability statement

The original contributions presented in this study are included in the article/**Supplementary material**, further inquiries can be directed to the corresponding author/s.

Ethics statement

The studies involving human participants were reviewed and approved by the Medical Committee

of Beijing Tongren Hospital, Capital Medical University (TRECKY2018-067). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

YJL, SZ, CZ, and SG: conceptualization. YJL, PC, and LY: data curation. JY: formal analysis. SZ and CZ: funding acquisition. LY, RR, and YL: methodology. SZ and DW: project administration. YJL, RR, and YL: visualization and writing-original draft. LY, DW, CZ, and SZ: writing-review and editing. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the National Natural Science Foundation of China (grant no. 81770989) and the Capital Health Research and Development of Special (grant no. 2020-2-2057) through SZ. It was also supported by the Beijing Postdoctoral Research Foundation (grant no. 2021-ZZ-007) and the China Postdoctoral Science Foundation (grant no. 2021M702310), through CZ.

Acknowledgments

We thank all the participants and their parents.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.973735/full#supplementary-material

References

Agterberg, M. J., Snik, A. F., Hol, M. K., van Esch, T. E., Cremers, C. W., Van Wanrooij, M. M., et al. (2011). Improved horizontal directional hearing in bone conduction device users with acquired unilateral conductive hearing loss. *J. Assoc. Res. Otolaryngol.* 12, 1–11. doi: 10.1007/s10162-010-0235-2

Agterberg, M. J., Snik, A. F., Hol, M. K., Van Wanrooij, M. M., and Van Opstal, A. J. (2012). Contribution of monaural and binaural cues to sound localization in listeners with acquired unilateral conductive hearing loss: Improved directional hearing with a bone-conduction device. *Hear. Res.* 286, 9–18. doi: 10.1016/j.heares.2012.02.012

Agterberg, M. J. H., Snik, A. F. M., Van de Goor, R. M. G., Hol, M. K. S., and Van Opstal, A. J. (2019). Sound-localization performance of patients with single-sided deafness is not improved when listening with a bone-conduction device. *Hear. Res.* 372, 62–68. doi: 10.1016/j.heares.2018.04.007

Bartel-Friedrich, S. (2015). Congenital auricular malformations: Description of anomalies and syndromes. *Facial Plast. Surg.* 31, 567–580. doi: 10.1055/s-0035-1568139

Dahm, V., Auinger, A. B., Liepins, R., Baumgartner, W. D., Riss, D., and Arnoldner, C. (2019). A randomized cross-over trial comparing a pressure-free, adhesive to a conventional bone conduction hearing device. *Otol. Neurotol.* 40, 571–577. doi: 10.1097/MAO.0000000000002184

Dobrev, I., Sim, J. H., Pfiffner, F., Huber, A. M., and Röösli, C. (2019). Experimental investigation of promontory motion and intracranial pressure following bone conduction: Stimulation site and coupling type dependence. *Hear. Res.* 378, 108–125. doi: 10.1016/j.heares.2019.03.005

Fu, X., Li, P., Zhang, L., Song, Y., An, Y., Zhang, A., et al. (2022). Activation of Rictor/mTORC2 signaling acts as a pivotal strategy to protect against sensorineural hearing loss. *Proc. Natl. Acad. Sci. U. S. A.* 119:e2107357119. doi: 10.1073/pnas.2107357119

Gao, Z., Wang, S., Yang, H., Feng, G., Shang, Y., Wang, B., et al. (2022). Cochlear implantation in young mandarin-speaking children: One year after first fitting. *Otol. Neurotol.* 43, e645–e650. doi: 10.1097/MAO.0000000000003555

He, Z. H., Li, M., Fang, Q. J., Liao, F. L., Zou, S. Y., Wu, X., et al. (2021). FOXG1 promotes aging inner ear hair cell survival through activation of the autophagy pathway. *Autophagy* 17, 4341–4362. doi: 10.1080/15548627.2021.1916194

Hong, G., Fu, X., Qi, J., Shao, B., Han, X., Fang, Y., et al. (2022). Dock4 is required for the maintenance of cochlear hair cells and hearing function. *Fundam. Res.* doi: 10.1016/j.fmre.2022.04.016

Jiang, Z., Wang, J., Feng, Y., Sun, D., Zhang, X., Shi, H., et al. (2021). Analysis of early biomarkers associated with noise-induced hearing loss among shipyard workers. *JAMA Netw. Open* 4:e2124100. doi: 10.1001/jamanetworkopen.2021.24100

Kunst, S. J., Leijendeckers, J. M., Mylanus, E. A., Hol, M. K., Snik, A. F., and Cremers, C. W. (2008). Bone-anchored hearing aid system application for unilateral congenital conductive hearing impairment: Audiometric results. *Otol. Neurotol.* 29, 2–7. doi: 10.1097/mao.0b013e31815ee29a

Liu, C. C., Livingstone, D., and Yunker, W. K. (2017). The role of bone conduction hearing aids in congenital unilateral hearing loss: A systematic review. *Int. J. Pediatr. Otorhinolaryngol.* 94, 45–51. doi: 10.1016/j.ijporl.2017.01.003

Liu, H., Zhang, H., Liu, S., Chen, X., Han, D., and Zhang, L. (2011). International outcome inventory for hearing aids (IOI-HA): Results from the Chinese version. *Int. J. Audiol.* 50, 673–678. doi: 10.3109/14992027.2011.588966

Luquetti, D. V., Heike, C. L., Hing, A. V., Cunningham, M. L., and Cox, T. C. (2012). Microtia: Epidemiology and genetics. *Am. J. Med. Genet. A* 158A, 124–139. doi: 10.1002/ajmg.a.34352

Nelissen, R. C., Agterberg, M. J., Hol, M. K., and Snik, A. F. (2016). Three-year experience with the Sophono in children with congenital conductive unilateral hearing loss: Tolerability, audiometry, and sound localization compared to a bone-anchored hearing aid. *Eur. Arch. Otorhinolaryngol.* 273, 3149–3156. doi: 10.1007/s00405-016-3908-6

Neumann, K., Thomas, J. P., Voelter, C., and Dazert, S. (2019). A new adhesive bone conduction hearing system effectively treats conductive hearing loss in children. *Int. J. Pediatr. Otorhinolaryngol.* 122, 117–125. doi: 10.1016/j.ijporl.2019.03.014

Okada, M., Welling, D. B., Liberman, M. C., and Maison, S. F. (2020). Chronic conductive hearing loss is associated with speech intelligibility deficits in patients with normal bone conduction thresholds. *Ear Hear.* 41, 500–507. doi: 10.1097/AUD.0000000000000787

Risoud, M., Hanson, J. N., Gauvrit, F., Renard, C., Lemesre, P. E., Bonne, N. X., et al. (2018). Sound source localization. *Eur. Ann. Otorhinolaryngol. Head Neck Dis.* 135, 259–264. doi: 10.1016/j.anorl.2018.04.009

Sabin, A. T., Macpherson, E. A., and Middlebrooks, J. C. (2005). Human sound localization at near-threshold levels. *Hear. Res.* 199, 124–134. doi: 10.1016/j.heares.2004.08.001

Shahar-Nissan, K., Oikawa Tepperberg, M., Mendelson, E., and Bilavsky, E. (2022). Retrospective identification of congenital cytomegalovirus infection using dried blood samples – missed opportunities and lessons. *J. Clin. Virol.* 152:105186. doi: 10.1016/j.jcv.2022.105186

Shirane, M., Ganaha, A., Nakashima, T., Shimoara, S., Yasunaga, T., Ichihara, S., et al. (2020). Comprehensive hearing care network for early identification and intervention in children with congenital and late-onset/acquired hearing loss: 8 years' experience in Miyazaki. *Int. J. Pediatr. Otorhinolaryngol.* 131:109881. doi: 10.1016/j.ijporl.2020.109881

Tao, Y., Liu, X., Yang, L., Chu, C., Tan, F., Yu, Z., et al. (2022). AAV-ie-K558R mediated cochlear gene therapy and hair cell regeneration. *Signal Transduct. Target. Ther.* 7:109. doi: 10.1038/s41392-022-00938-8

Thompson, N. J., Kane, S. L. G., Corbin, N. E., Canfarotta, M. W., and Buss, E. (2020). Spatial hearing as a function of presentation level in moderate-to-severe unilateral conductive hearing loss. *Otol. Neurotol.* 41, 167–172. doi: 10.1097/MAO.0000000000002475

Tillein, J., Hubka, P., and Kral, A. (2016). Monaural congenital deafness affects aural dominance and degrades binaural processing. *Cereb. Cortex* 26, 1762–1777. doi: 10.1093/cercor/bhv351

Van Wanrooij, M. M., and Van Opstal, A. J. (2004). Contribution of head shadow and pinna cues to chronic monaural sound localization. *J. Neurosci.* 24, 4163–4171. doi: 10.1523/JNEUROSCI.0048-04.2004

van Wieringen, A., Boudewyns, A., Sangen, A., Wouters, J., and Desloovere, C. (2019). Unilateral congenital hearing loss in children: Challenges and potentials. *Hear. Res.* 372, 29–41. doi: 10.1016/j.heares.2018.01.010

Vogt, K., Frenzel, H., Ausili, S. A., Hollfelder, D., Wollenberg, B., Snik, A. F. M., et al. (2018). Improved directional hearing of children with congenital unilateral conductive hearing loss implanted with an active bone-conduction implant or an active middle ear implant. *Hear. Res.* 370, 238–247. doi: 10.1016/j.heares.2018.08.006

Vogt, K., Wasmann, J. W., Van Opstal, A. J., Snik, A. F. M., and Agterberg, M. J. H. (2020). Contribution of spectral pinna cues for sound localization in children with congenital unilateral conductive hearing loss after hearing rehabilitation. *Hear. Res.* 385:107847. doi: 10.1016/j.heares.2019.107847

Vyskocil, E., Liepins, R., Kaider, A., Blineder, M., and Hamzavi, S. (2017). Sound localization in patients with congenital unilateral conductive hearing loss with a transcutaneous bone conduction implant. *Otol. Neurotol.* 38, 318–324. doi: 10.1097/MAO.0000000000001328

Wang, M., Dong, Y., Gao, S., Zhong, Z., Cheng, C., Qiang, R., et al. (2022). Hippo/YAP signaling pathway protects against neomycin-induced hair cell damage in the mouse cochlea. *Cell. Mol. Life Sci.* 79:79. doi: 10.1007/s00018-021-04029-9

Wang, S., Mannell, R., Newall, P., Zhang, H., and Han, D. (2007). Development and evaluation of Mandarin disyllabic materials for speech audiometry in China. *Int. J. Audiol.* 46, 719–731. doi: 10.1080/14992020701558511

Weiss, R., Leinung, M., Baumann, U., Weißgerber, T., Rader, T., and Stöver, T. (2017). Improvement of speech perception in quiet and in noise without decreasing localization abilities with the bone conduction device Bonebridge. *Eur. Arch. Otorhinolaryngol.* 274, 2107–2115. doi: 10.1007/s00405-016-4434-2

Wood, K. C., Town, S. M., and Bizley, J. K. (2019). Neurons in primary auditory cortex represent sound source location in a cue-invariant manner. *Nat. Commun.* 10:3019. doi: 10.1038/s41467-019-10868-9

World Health Organization [WHO] (2021). *Addressing the Rising Prevalence of Hearing Loss*. Available online at: https://apps.who.int/iris/bitstream/handle/10665/260336/9789241550260-eng.pdf?fbclid=IwAR0tQbo1bvObm8h5hD0PBPmfPowiHLiQIDAJqJcnV1ThYnDy1fnb-q9flc (accessed February 20, 2021).

Yu, J. K., Wong, L. L., Tsang, W. S., and Tong, M. C. (2014). A tutorial on implantable hearing amplification options for adults with unilateral microtia and atresia. *Biomed. Res. Int.* 2014:703256. doi: 10.1155/2014/703256

Zhang, Y., Fang, Q., Wang, H., Qi, J., Sun, S., Liao, M., et al. (2022). Increased mitophagy protects cochlear hair cells from aminoglycoside-induced damage. *Autophagy* 1–17. [Epub ahead of print]. doi: 10.1080/15548627.2022.2062872

frontiers | Frontiers in Neuroscience

TYPE Technology and Code
PUBLISHED 01 September 2022
DOI 10.3389/fnins.2022.904003

OPEN ACCESS

EDITED BY
Preben Kidmose,
Aarhus University, Denmark

REVIEWED BY
Han-Jeong Hwang,
Korea University, South Korea
Walter Besio,
University of Rhode Island,
United States

*CORRESPONDENCE
Steffen Dasenbrock
steffen.dasenbrock@uol.de

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

RECEIVED 25 March 2022
ACCEPTED 05 August 2022
PUBLISHED 01 September 2022

CITATION
Dasenbrock S, Blum S, Maanen P, Debener S, Hohmann V and Kayser H (2022) Synchronization of ear-EEG and audio streams in a portable research hearing device.
Front. Neurosci. 16:904003.
doi: 10.3389/fnins.2022.904003

Synchronization of ear-EEG and audio streams in a portable research hearing device

Steffen Dasenbrock[1,2]*, Sarah Blum[2,3], Paul Maanen[2,3], Stefan Debener[2,3], Volker Hohmann[1,2] and Hendrik Kayser[1,2]

[1]Auditory Signal Processing and Hearing Devices, Department of Medical Physics and Acoustics, University of Oldenburg, Oldenburg, Germany, [2]Cluster of Excellence "Hearing4all", University of Oldenburg, Oldenburg, Germany, [3]Neuropsychology Lab, Department of Psychology, University of Oldenburg, Oldenburg, Germany

Recent advancements in neuroscientific research and miniaturized ear-electroencephalography (EEG) technologies have led to the idea of employing brain signals as additional input to hearing aid algorithms. The information acquired through EEG could potentially be used to control the audio signal processing of the hearing aid or to monitor communication-related physiological factors. In previous work, we implemented a research platform to develop methods that utilize EEG in combination with a hearing device. The setup combines currently available mobile EEG hardware and the so-called Portable Hearing Laboratory (PHL), which can fully replicate a complete hearing aid. Audio and EEG data are synchronized using the Lab Streaming Layer (LSL) framework. In this study, we evaluated the setup in three scenarios focusing particularly on the alignment of audio and EEG data. In Scenario I, we measured the latency between software event markers and actual audio playback of the PHL. In Scenario II, we measured the latency between an analog input signal and the sampled data stream of the EEG system. In Scenario III, we measured the latency in the whole setup as it would be used in a real EEG experiment. The results of Scenario I showed a jitter (standard deviation of trial latencies) of below 0.1 ms. The jitter in Scenarios II and III was around 3 ms in both cases. The results suggest that the increased jitter compared to Scenario I can be attributed to the EEG system. Overall, the findings show that the measurement setup can time-accurately present acoustic stimuli while generating LSL data streams over multiple hours of playback. Further, the setup can capture the audio and EEG LSL streams with sufficient temporal accuracy to extract event-related potentials from EEG signals. We conclude that our setup is suitable for studying closed-loop EEG & audio applications for future hearing aids.

KEYWORDS

hearing aids, mobile EEG, portable setup, timing, jitter, ear-EEG, cEEGrid, neuro-steered hearing device

1. Introduction

Current neuroscientific discoveries have led to the concept of using brain signals as additional input to control audio signal processing in hearing devices (Slaney et al., 2020). In this process, correlates from the electroencephalography (EEG) recording are used to make statements about the user's mental state, such as auditory attention (O'sullivan et al., 2015) or listening effort (Bernarding et al., 2012; Haro et al., 2022). These can potentially be used to adjust the signal processing in the hearing device, for example, a beamformer that amplifies the attended speaker while suppressing the ignored ones (Aroudi and Doclo, 2019). This idea is commonly referred to as neuro-steered hearing devices, cognitively controlled hearing aids, or similar (O'Sullivan et al., 2017; Das et al., 2020; Geirnaert et al., 2021). Current developments in miniaturization and improving the wearability of (ear-centered) EEG suggest that the relevant neural signals can still be detected when using fewer electrodes and less spatial coverage compared to traditional cap EEG setups (Kidmose et al., 2012; Debener et al., 2015; Mikkelsen et al., 2015; Bleichner et al., 2016; Fiedler et al., 2017). This research is relevant to realize closed-loop applications on wearable devices with high wearing comfort.

To develop EEG-based hearing aid signal processing methods, a portable platform is required that offers the signal quality and timing accuracy required for closed-loop applications with audio and EEG signals. In Dasenbrock et al. (2021), we introduced a research platform, which combines the mobile Smarting EEG system and the so-called Portable Hearing Laboratory (PHL, Pavlovic et al., 2018). The PHL is a portable research hearing device that can fully replicate a hearing device. It uses the open-source open Master Hearing Aid (openMHA, Kayser et al., 2022) software for real-time, low-latency audio signal processing. The EEG data is received over a wireless network connection. The Lab Streaming Layer (LSL, Kothe et al., 2014) is used to synchronize audio and EEG data. LSL is a framework that can synchronize data streams from different devices by measuring their respective clock drift over a network connection to map the locally generated time stamps into one common timeline. First pilot timing and physiological tests showed that this setup's portable components could be used to extract event-related potentials (ERPs) using an Oddball paradigm (Dasenbrock et al., 2021).

The temporal synchronicity of audio and EEG data is crucial for such a setup. Alignment inaccuracies of audio and EEG data may lead to errors in the data analysis that may affect the predictive power of the following listening state analysis. Several sources of inaccuracy exist when aligning audio and EEG data. For example, there may be variations in the playback of the stimuli, i.e., differences between the software event markers and the actual playback of the device. Further, inaccuracies can occur when creating an LSL stream of continuous time-series data, such as the EEG. Generally, factors such as the operating system, drivers, and hardware performance will typically introduce variations in delays; thus, a latency variation (jitter) is always expected in real systems. Timing accuracy, especially stability (between sessions), must also be considered for online applications, which rely on a constant latency between EEG and audio.

Time synchronization has long been a challenge in the implementation of mobile EEG systems. When performing EEG experiments, the amount of jitter needs to be sufficiently small. The required temporal precision generally depends on the method used. It is particularly relevant for investigating time-averaged data. For instance, obtaining ERPs requires the extraction of EEG trials by epoching the data using event markers, which indicate a response-evoking feature of the auditory stream. Accurate time synchronization is crucial since it leads to an exact alignment when averaging over the single-trial responses, leading to high and sharp ERP components, i.e., specific peak amplitudes in the ERP (Williams et al., 2021). For wireless commercial mobile EEG systems, the problem of event-locking the EEG data has long been challenging, as they are often not designed for it. For instance, in early iterations of the wireless Emotiv EEG system, it was found that the built-in event-locking was unstable and did not produce high-quality ERPs (Hairston et al., 2014; Ries et al., 2014). Nowadays, mobile EEG hardware can also be used for ERP studies, as indicated by recent studies of the Emotiv EEG system (Williams et al., 2021). Further, fully mobile smartphone-based systems have already been successfully coupled with mobile Smarting EEG systems (mBrainTrain, Belgrade, Serbia), which could be used for extracting ERPs outside the laboratory (Debener et al., 2015; Blum et al., 2017; Hölle et al., 2022). Building upon the smartphone-based approach, the setup employed in this work focuses on hearing aid applications, leveraging the real-time capabilities of the PHL. The setup presented by our group in Dasenbrock et al. (2021) was extended by the possibility of sending single software event markers, which enabled us to perform a comprehensive timing analysis to evaluate its suitability for research into closed-loop hearing devices with EEG.

The timing precision of the setup was systematically evaluated in several timing test scenarios, that address the different components of the system. As the setup is composed of two completely independent components, namely the PHL and the Smarting EEG system, the timing precision of the setup is assumed to be composed of the timing accuracy of the PHL and the timing accuracy of the EEG system. The accuracy of these two systems was examined separately in timing test scenarios I and II. Timing test scenario I examines the PHL's ability to create a precise software marker that marks the playback time of an acoustic stimulus, i.e., the event marker. In this test, the event marker time is compared to the actual audio playback of the PHL. Timing test scenario II examines how precisely the EEG system samples and time stamps an

incoming signal by comparing the sampled and time stamped signal from the EEG system against a known reference. A third test, i.e., timing test Scenario III, was conducted to measure the whole system's timing, as it would be present in a real EEG measurement. In this final test, a reference signal generated by the PHL is compared to the sampled and time stamped signal of the EEG system. To investigate differences in the timing accuracy between shorter and longer durations, all timing tests were performed for 15 min and 3 h. No human subjects were used in the tests, as their introduction would lead to additional between-subject variability (Intriligator and Polich, 1995), as well as measurement variability (Callaway and Halliday, 1973). Thus, this study was kept purely technical, focusing on the effects of the measurement devices.

2. Materials and methods

In the following two sections, the setup and its technical features as well as the timing tests are described in detail.

2.1. Setup

This section describes the technical aspects of the setup. For this purpose, the two subsystems, i.e., the PHL and the EEG system, are discussed in more detail. In addition, the LSL framework is discussed in depth, followed by a general system description. Figure 1 outlines how the setup is carried in a measurement.

2.1.1. Portable hearing laboratory

The PHL is the central component of the setup. It is an integrated hearing aid research platform combining portable hardware and the openMHA. Due to its portability, the device is particularly suitable for field studies to investigate more realistic hearing scenarios outside the laboratory. A personal computer or smartphone controls the PHL through a WiFi connection. It consists of a main unit, including a battery, and supports several ear-level transducers. The setup uses a binaural 4-microphone behind-the-ear (BTE) hearing aid headset developed for the PHL. As visualized in Figure 1, a strap connected to the back of the device is used to carry the PHL around the neck, which weights about 130 g with dimensions of 58 × 90 × 30 mm. The PHL runs a Linux operating system (OS)–MAHALIA (Obbard and James, 2018), optimized for the device's hardware and to run openMHA. The PHL supports sampling rates in the range of 8 and 96 kHz. In this work, a sampling rate of 16 kHz was used. Figure 2 highlights some key features of the PHL's software and hardware. An extended table with detailed technical information on the PHL can be found in the Supplementary material.

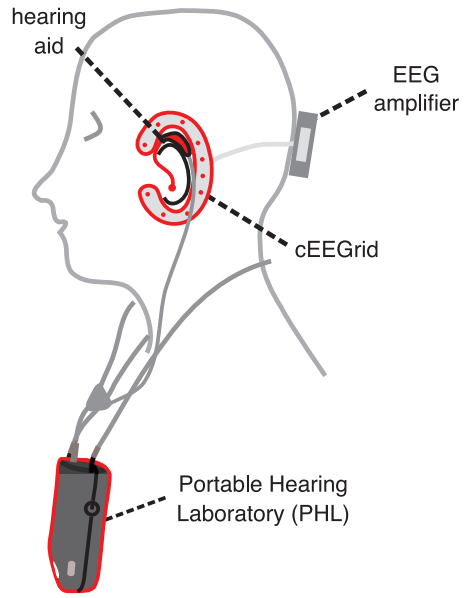


FIGURE 1
Sketch of the hearing aid and EEG setup carried in a measurement. A strap connected to the back of the device is used to carry the Portable Hearing Laboratory (PHL) around the neck. Acoustic stimuli are presented *via* the hearing aids connected to the PHL. The around-the-ear EEG sensor cEEGrid is used to measure the neural activity. The cEEGrid is connected to the mobile wireless Smarting EEG amplifier. Both audio and EEG data streams are captured on the PHL for further processing and recording. Figure adapted from Dasenbrock et al. (2021).

Hardware	Software
• weight: 130 g • connector to 4-mic BTE-heading aid headset	• OS: MAHLIA (Linux) • runs openMHA signal processing software

FIGURE 2
Photo of Portable Hearing Laboratory (PHL) and table with selection of hardware and software features. The hardware consists of a portable main unit and a binaural 4-microphone behind-the-ear (BTE) hearing aid headset. Photo adapted from Kayser et al. (2022).

openMHA is an open-source software platform for real-time, low-latency hearing aid signal processing (Kayser et al., 2022). The software is implemented in C++ and contains a

wide range of audio processing algorithms. Its modular structure allows adding new functionalities in the form of plugins. The functions of a specific openMHA instance (such as algorithms, inputs, outputs, and sampling rate) are defined by the user using the openMHA configuration language consisting of line-based human-readable text commands. The software runs on several computer systems, including lab and portable setups, particularly the PHL.[1]

2.1.2. EEG system

The EEG system used in the presented setup comprises three parts, i.e., (1) electrode setup, (2) Smarting EEG amplifier, and (3) LSL streaming interface. The setup aims to be used mainly with cEEGrid electrodes (Debener et al., 2015; Bleichner and Debener, 2017). The cEEGrid is a disposable electrode grid with printed sensor arrays based on flex-print technology. Ten circular electrodes are arranged in a c-shape to fit around the ear and are attached using adhesive tape. The electrode setup is connected to a small, wireless 24 channel SMARTING EEG amplifier (mBrainTrain, Belgrade, Serbia), placed at the back of the head. The amplifier receives the EEG signal, amplifies it, and streams a Bluetooth signal which is received by the Smarting Android application on the smartphone. The data transfer and synchronization between the Smarting amplifier and the smartphone is handled by a proprietary protocol unknown to the authors. The smartphone and its Smarting app act as the LSL streaming interface. It streams the signal over the network using the LSL framework. In this work, the SMARTING EEG amplifier (Serial number: 010016) is used in combination with a Sony Xperia Z1 smartphone (model: C6903; OS: Android 5.1.1) using a pre-installed Smarting Android application (Version: 1.6.0). A sampling rate of 250 Hz was used. More information on the EEG system can be found in the Supplementary material.

2.1.3. Lab streaming layer

When dealing with multiple components within one setup, each device typically relies on its built-in high-resolution clock. Even if the time difference Δt_{clocks} between the clocks is known at some time, Δt_{clocks} usually changes over time due to clock drift caused by clocks counting at slightly different rates. One possible way of synchronizing data streams of multiple sources is to use hardware synchronization, usually achieved using TTL (transistor-transistor logic) signals, e.g., by regularly sending out synchronization pulses to the attached devices (Reis et al., 2014). However, this usually requires wiring up the different devices. Repeated synchronization of data streams can also be achieved without the need for wires and connectors by using (wireless) networks and a software agent running on each device. For this purpose, the LSL was used. LSL is an open framework that consists of a core library, interfaces for many common programming languages, and several tools (Kothe et al., 2014; Blum et al., 2021). It can be used to measure time differences between connected devices in a network-based setup. Thus, recording setups including LSL can consist of several pieces of hardware and software. LSL tools include a recording program, i.e., the LabRecorder, file importers, and apps to support many EEG systems on the market. The website offers extensive documentation on LSL's functionalities and tools and provides a list of the many different pieces of hardware which have adopted the LSL standard.[2] The built-in time synchronization capability of LSL is designed for sub-millisecond accuracy on a network of computers connected *via* WiFi.

In LSL terminology, the combination of the raw data from a device and its metadata, such as channel count or sampling rate, is referred to as a stream. LSL streams can have a regular sampling rate, such as continuous EEG, or an irregular sampling rate, such as event markers, e.g., when marking an acoustic event, such as the onset of a sound stimulus. The time synchronization of LSL relies on two pieces of data being collected in addition to the actual sample data: (1) timestamp, (2) clock correction offset. For each LSL sample, a timestamp is read from a local high-resolution clock of the device. The clock correction offset is a measurement of the momentary offset between the two involved clocks and is computed at periodic intervals, by default, every 5 s. LSL is not limited to the use of only two devices. If multiple devices are present in the setup, the clock correction offset between the receiver and every sender is measured.

LSL uses a protocol similar to the Network Time Protocol to measure the clock correction offset. The simplest way to map the time series data from different devices into a common timeline is to add the most recent clock correction offset value to each remotely collected timestamp. Other more sophisticated methods exist that attempt to smooth the clock correction values, such as an outlier-resistant (robust) linear fit through a history of clock correction offsets to reduce the effects of jitter in the clock correction offset measurement. Further, after applying the clock correction offsets, there is a second source of jitter, i.e., jitter in the time stamps. This jitter is not due to synchronization but because time-sampling is usually not done at regular intervals but on a slightly stochastic schedule (determined by the hardware, driver, and operating system). If the LSL stream has a regular sampling rate, this jitter can also be reduced by applying smoothing algorithms. In this work,

1 In this study, the openMHA 4.17.0 release version was used, which is contained in the corresponding MAHALIA image 4.17.0.; MAHALIA download: http://mahalia.openmha.org (accessed March 18, 2022).

2 More information on the LabRecorder, file importers, and apps can be found in the LSL documentation https://labstreaminglayer.readthedocs.io/index.html (accessed June 25, 2022).

however, no smoothing of the clock correction offsets and time stamps was applied.

In the setup presented in this work, the LSL framework was used for the synchronization of EEG data and audio event markers. A network link is established by the smartphone's connection (which runs the Smarting app) to the WiFi hotspot provided by the PHL. The possibility of using LSL in the presented setup is provided on the one hand by the Smarting app's functionality to generate LSL streams of the EEG data, and on the other hand by openMHA's interface to LSL. Generally, to use LSL, all clients in the network need to support LSL. Open source Android application projects for LSL streaming and recording enable the inclusion of additional sensor streams (Blum et al., 2021).

2.1.4. System description

Figure 3A schematically illustrates which components of the setup are responsible for the different signals and data streams. A total of four measurement points, a-d, were included in the drawing to illustrate between which points the timing was measured. The measurement points will be relevant in Section 2.2, which provides a detailed description of the timing test scenarios.

The setup's data flow is described in the scheme of a sender-receiver architecture running on the PHL (left) and the EEG system (right). The purpose of the sender instance (top left) is to provide acoustic stimuli *audio out* to the subject *via* the hearing aids. The corresponding physical voltage signal is measured at measurement point a. The sender instance simultaneously generates an *audio event marker LSL stream* containing event markers that specify the beginning of a stimulus onset in the audio signal. Point d denotes the position in the data flow at which the *audio event marker LSL stream* is measured.

At measurement point b, the resulting EEG voltage signal (e.g., captured with cEEGrid electrodes) is fed into the EEG amplifier. The smartphone receives the EEG data from the amplifier *via* a Bluetooth connection and creates an LSL stream of the EEG data, referred to as EEG-LSL stream. The *EEG LSL stream* is measured at point c.

The smartphone and the PHL share the same network. This allows the LSL framework to collect all necessary information to map both streams into a common timeline (see Section 2.1.3 for details). Both the *audio event marker LSL stream* and the *EEG LSL stream* are captured by the receiver instance (bottom left). Both sender and receiver instances were implemented as two independent openMHA instances with different processing configurations that run on the PHL.[3]

3 The openMHA configurations for both sender and receiver instances are published and can be found on Github: https://github.com/steffendasenbrock/SynchronizationEarEEGAudioStreams

2.2. Timing tests

This section describes the different timing test scenarios performed to investigate the timing properties of the setup outlined in the previous section. It is important to note that the timing tests were performed without humans and an electrode setup, such as the cEEGrid. The signals fed into the EEG system were routed directly into the EEG amplifier. The audio signals from the PHL were also not played directly through the hearing aids but redirected to an audio jack. In each scenario, the timing was evaluated between two of the four measurement points a-d (see Figure 3A). The timing tests of the different scenarios were performed separately, i.e., two different measurement points were measured simultaneously in each scenario.

We conducted tests in three different scenarios. The different scenarios are discussed in more detail in the following sections. An additional test described in Section 2.2.3.1 was performed to compare the recording capabilities of the receiver instance (Figure 3A, bottom left) with an established reference recording software.

Stimuli

For all tests, a rectangular pulse was used as a test signal to investigate the time synchronization properties of the setup, as it produces sharp detectable responses. Sixty millisecond pulses were repeated at 1 Hz. All timing test scenarios were performed for two different durations, i.e., 15 min (short) and 3 h (long). For the short timing test 900 and the long timing test, 10,800 trial latencies were determined. A total of 21 timing tests were performed.

2.2.1. Scenario I: Sender instance timing

In timing test Scenario I, the timing accuracy of the sender instance (Figure 3A, top left) was evaluated. The sender instance was programmed to play the test stimulus containing the rectangular pulses. At the same time, it was configured to generate an LSL event marker whenever a rising edge is detected in the signal. More details on the implementation of this mechanism are described in Supplementary Figure 2. In Figure 3B (left), the procedure to test the sender instance's accuracy is sketched in a timing diagram. To specify the timing of the sender instance, the latency Δt_n was measured. In this scenario, Δt_n was defined as the time difference between the timestamps of the *audio event marker LSL stream* at measurement point d and the rising edges in the actual playback signal *audio out* at measurement point a.

The LabStreamer (NeuroBehavioral Systems; Albany, CA, USA) was used to measure Δt_n. The LabStreamer is a commercial device that can be considered an oscilloscope optimized for network timing, designed explicitly for analyzing timing precision when dealing with the LSL framework. It provides a sampling rate of 10 kHz, which

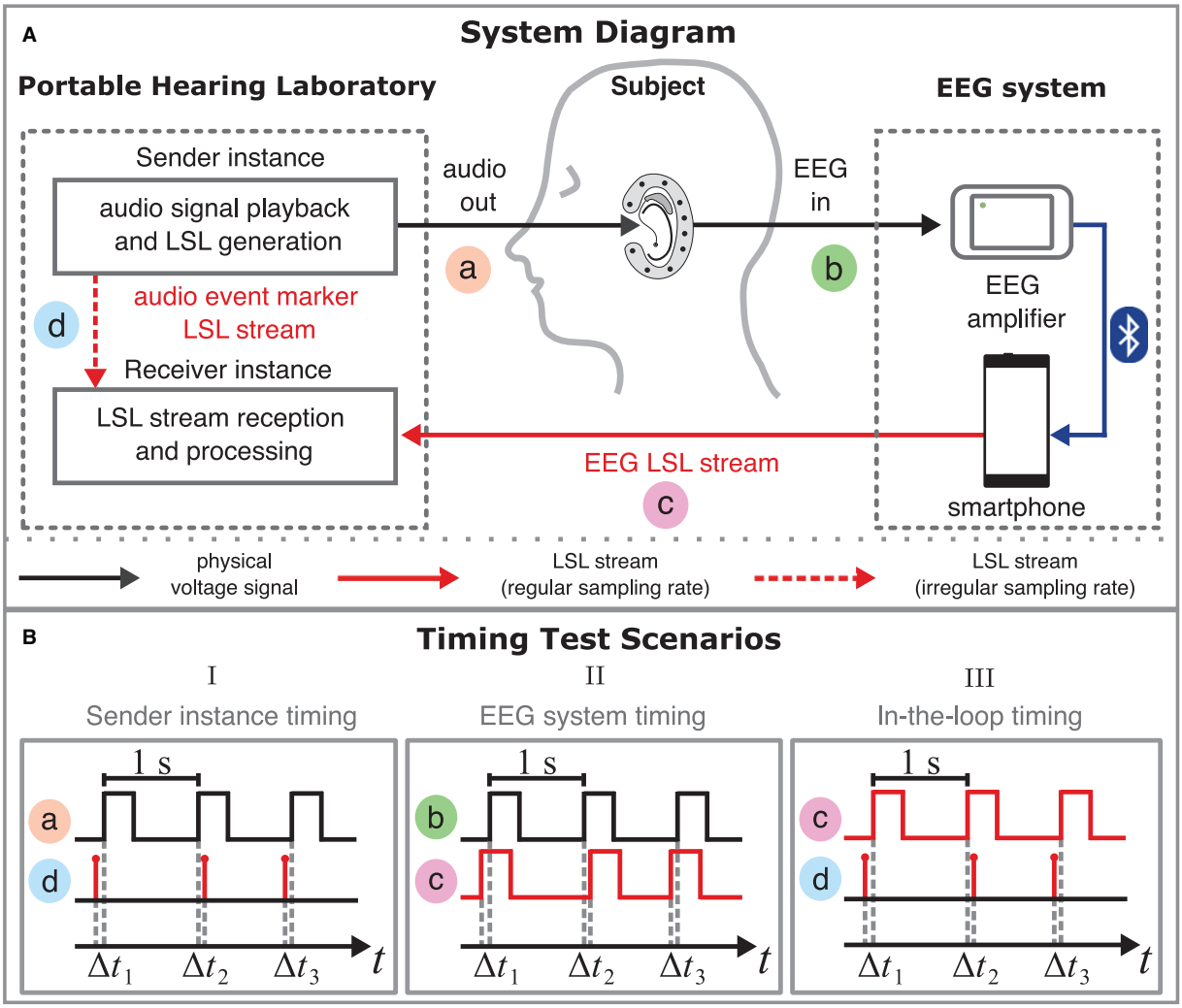


FIGURE 3
(A) System diagram of the measurement setup. The system diagram shows the signal flow and illustrates which components of the setup are responsible for the different signals and data streams and how they are related. Black lines refer to physical voltage signals; red lines refer to LSL streams. The setup combines the Portable Hearing Laboratory (PHL, left) and the EEG system (right). The PHL's function is described in the scheme of a sender-receiver architecture. The sender instance's (top left) role is to present acoustic stimuli *audio out* to the subject *via* the hearing aids. The physical voltage signal is measured at measurement point a. During playback of the stimuli, the sender instance simultaneously creates an *audio event marker LSL stream* that contains event markers indicating specific time points in the audio signal. The *audio event marker LSL stream* is measured at point d. Subsequently, the resulting EEG voltage signal *EEG in* (measurement point b) is amplified using the mobile EEG amplifier. The smartphone receives the EEG data *via* a Bluetooth connection and creates an LSL stream of the EEG data, i.e., *EEG LSL stream* (measurement point c). The receiver instance captures both the *audio event marker LSL stream* and the *EEG LSL stream*. **(B)** Timing diagrams for all three timing test scenarios. Timing diagrams relate two signals or streams in terms of time (x-axis). Square wave signals were used to test the timing in the setup. The time difference between two related time points is defined as trial latencies Δt_n, measured about every second. In timing test Scenario I (left) Δt_n was computed by comparing the rising edges in the voltage signal *audio out* (a) and the *audio event marker LSL stream* (d). In timing test Scenario II (center) Δt_n was computed by comparing the rising edges in the voltage signal *EEG in* (b) and the *EEG LSL stream* (c). In timing test Scenario III (right) Δt_n was computed by comparing the rising edges in the *EEG LSL stream* (c) and the *audio event marker LSL stream* (d).

results in a measuring accuracy of 0.1 ms. It features different input and output channels such as audio, general analog inputs and outputs, and the ability to receive and generate LSL streams. It offers an oscilloscope panel to trigger signals by LSL events and a latency histogram panel. More detailed information on the LabStreamer can be found in the Supplementary material and in the respective documentation.[4]

4 Neurobehavioral Systems. LabStreamer https://www.neurobs.com/menu_presentation/menu_hardware/labstreamer (accessed June 24, 2022).

For the technical implementation, a network connection was established between PHL and LabStreamer to use LSL. The analog signal of the PHL was fed directly into the audio input jack of the LabStreamer. More technical details can be found in Supplementary Section 1.1 of the Supplementary material.

2.2.2. Scenario II: EEG system timing

Timing test Scenario II was done to investigate the timing accuracy of the EEG system (Figure 3A, right) when it converts an analog voltage signal into an LSL stream. Figure 3B (center) visually describes the procedure to test the EEG system's accuracy in a timing diagram. In this scenario, the latency Δt_n was defined as the time difference between the rising edges in the *EEG LSL stream* at measurement point b and the rising edges contained in the analog *EEG in* test signal fed into the EEG system at measurement point c.

As done in the first scenario, the LabStreamer was used to measure Δt_n. For the technical implementation, a network connection was established between the EEG system and LabStreamer to use LSL, as in timing test Scenario I. The LabStreamer was used to feed the analog test signal into the EEG system and to capture the *EEG LSL stream*. More technical details about this procedure can be found in Supplementary Section 1.2 of the Supplementary material.

2.2.3. Scenario III: In-the-loop timing

In Scenarios I and II, the timing accuracy of the sender instance and EEG system were tested separately. In timing test Scenario III, described in this section, the timing accuracy of the whole setup was measured. For this, the *audio event marker LSL stream* of the sender instance was related to the *EEG LSL stream* of the EEG system. Figure 3B (right) outlines the procedure to test the whole setup's accuracy in a timing diagram. The latency Δt_n was defined here as the time difference between the timestamps of the sender instance's *audio event marker LSL stream* at measurement point d and the rising edges in the *EEG LSL stream* at measurement point c.

Timing test Scenario III was technically realized by feeding the *audio out* signal of the PHL's sender instance directly into the EEG system as *EEG in* signal (see Figure 3). This approach was adapted from Blum et al. (2017). Further details can be found in Supplementary Section 1.3 of the Supplementary material. In contrast to Scenarios I and II, the LabStreamer was not used here. Both *audio event marker LSL stream* and *EEG LSL stream* were recorded using the receiver instance on the PHL.

The calculation of Δt_n was done in a post-analysis. The EEG signal was interpolated sample-wise to correspond to the audio signal's sampling rate to determine the rising edge position. This model assumption is justified by the properties of the test signal, which features vertical edges. The EEG signal was epoched from −100 to 150 ms and baseline corrected from −100 to −50 ms with reference to the timestamps in the *audio event marker LSL stream*. As done in Blum et al. (2017), the latency Δt_n for each rectangular pulse was determined by calculating the time difference between the event marker time and the time when the EEG signal amplitude exceeded the half-maximum of the trial averaged response.

Metrics

The lag describes the arithmetic mean within a timing test session, i.e.,

$$\text{lag} = \frac{1}{N}\sum_{n=1}^{N}\Delta t_n, \tag{1}$$

and the jitter the standard deviation of the latency within a timing test session, i.e.,

$$\text{jitter} = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N} \mid \Delta t_n - \text{lag} \mid^2}. \tag{2}$$

An across-session range ΔR was calculated to quantify the spread of lag and jitter between timing tests of the same condition. ΔR was defined here as the difference between the maximum and the minimum value of lag and jitter across all timing tests of the same condition.

2.2.3.1. Comparison with reference data recording

Two additional 3 h in-the-loop recordings were conducted to compare the data received by the receiver instance with the data received by the standard recording program LabRecorder.[5] For this, LabRecorder and the receiver instance were used simultaneously on the PHL. The raw LSL data consisting of time series, time stamps, and clock corrections collected from the receiver instance and LabRecorder were compared to check if the receiver instance correctly recorded the data.

3. Results

In the following, the results of the previously described timing test scenarios are reported separately. Table 1 gives an overview of the obtained lag, jitter, and the across-session range ΔR of the lag and jitter for each condition.

3.1. Scenario I: Sender instance timing

Figure 4I shows the latency between the LSL timestamps of the *audio event marker LSL stream* and the rising edges in the actual playback signal of the PHL. The Lag was around 32.7 ms for both short and long durations and differed up to 0.14 ms

5 LSL. LabRecorder. https://github.com/labstreaminglayer/App-LabRecorder (accessed February 3, 2022).

TABLE 1 Measurement results in terms of lag, jitter, and across-session range ΔR for all three timing test scenarios.

Duration	Meas. Number	Lag in ms			Jitter in ms		
		I	II	III	I	II	III
15 min	1	32.69	−1.56	29.75	0.07	3.06	1.22
	2	32.76	−53.07	24.92	0.09	3.91	1.5
	3	32.81	−10.5	37.98	0.07	3.02	1.49
	4	32.68	−1.98	29.8	0.09	2.05	1.41
	5	32.67	−21.67	15.36	0.07	1.84	1.24
	ΔR	0.14	51.51	22.62	0.02	2.07	0.28
3 h	1	32.64	−13.2	25.19	0.09	3.82	3.33
	2	32.65	−7.89	24.61	0.09	2.84	2.99
	ΔR	0.01	5.31	0.58	0	0.98	0.34

ΔR refers to the difference between the maximum and the minimum value of lag and jitter across all timing tests within one scenario and duration. The columns labeled with Roman numerals belong to the respective timing test Scenarios I–III (see Sections 2.2.1–2.2.3). Five measurements were performed in the 15 min condition **(top)**; two measurements were performed in the 3 h condition **(bottom)**.

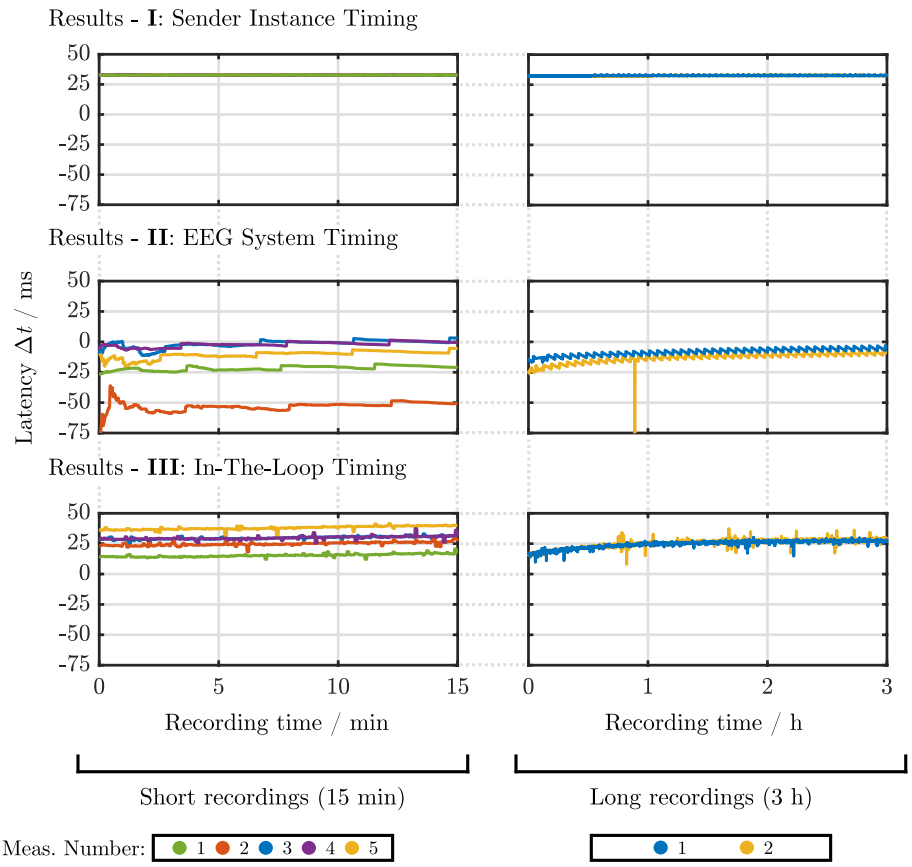


FIGURE 4
Latency-recording time curves for all timing test scenarios. The different plots show the course of the measured latency (y-axis) in milliseconds over recording time (x-axis) in minutes **(left)** or hours **(right)**. Five measurement runs were performed in the 15 min condition (left column); two measurement runs were performed in the 3 h condition (right column), denoted by different colors. The upper row shows the results for timing test Scenario I: sender instance timing (see Section 2.2.1); the middle row shows the results for timing test Scenario II: EEG system timing (see Section 2.2.2), and the bottom row shows the results for timing test Scenario III: In-the-loop timing (Section 2.2.3).

from session to session. Jitters were below 0.1 ms. It can be observed that the latency stayed constant over time for both 15 min and 3 h playback time, i.e., no temporal latency drift was observed.

3.2. Scenario II: EEG system timing

Figure 4II presents the latency plotted over recording time between rising edges in the *EEG LSL stream* and the rising edges contained in the analog test signal fed into an EEG system. Contrary to Scenario I the lag did not stay constant, but changed from session to session, indicated by a ΔR of up to 52 ms for the 15 min recording condition. Jitters were, however, relatively constant between sessions, ranging from 1.84 to 3.91 ms and 2.84 to 3.82 ms for the short and long timing test duration. A negative lag was observed in this scenario, which is further discussed in Section 4. Further, a positive, non-monotonic drift can be seen in both cases (short and long timing test duration), i.e., an increasing latency over time. For the 15 min timing test, after substantial fluctuations in the initial 200 s, the latency increases with an approximately linear trend. However, this linear trend no longer persists in the 3 h timing tests, and the latency flattens over time.

3.3. Scenario III: In-the-loop timing

Figure 4III shows the latency obtained from the in-the-loop system, i.e., the latency between the timestamps of the sender instance's *audio event marker LSL stream* and the rising edges in the *EEG LSL stream*. Similar to Scenario II lags differed from measurement to measurement with a ΔR of up to 22 ms. Jitters stayed relatively constant within each recording condition, ranging from 1.24 to 1.5 ms and from 2.99 to 3.33 ms for the 15 min and 3 h timing tests, respectively. The latency-recording time plots feature a similar trend as observed in Scenario II, as the recordings show a linear latency drift for the 15 min timing tests, which flattens out over time, visible in the 3 h timing tests.

3.3.1. Comparison with reference data recording

The comparison of receiver instance and LabRecorder data yielded two main results: (1) The raw LSL time series and the locally (in the EEG system and PHL's sender instance) created LSL time stamps were identical for both recording methods. (2) Comparing the LSL clock correction offset values obtained with openMHA and LabRecorder, differences occurred between the two recording methods. However, the long term average amounted to below 0.1 ms for both 3 h measurements, sporadic outliers up to 150 ms occurred for single samples. Figure 5 shows a histogram describing the relative occurrences of clock differences. For illustration purposes outlier values outside of $\pm$ 10 ms were included in the outer 10 ms bins.

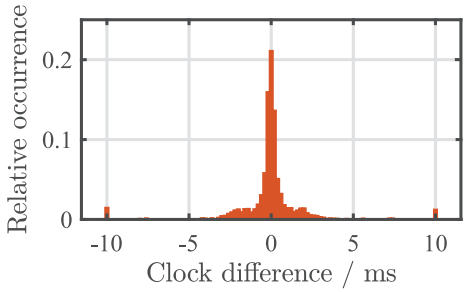


FIGURE 5
Histogram of one of the two 3 h recordings showing the differences between clock correction offset values obtained using openMHA's receiver instance and LabRecorder. Outlier values outside of $\pm$ 10 ms were included in the outer 10 ms bins.

The slight differences in the obtained clock correction values are negligible on average and can be explained by different points in time at which openMHA and LabRecorder request the clock correction values from the LSL framework. More prominent outliers of the clock correction can usually be eliminated by an outlier resilient fit method of the most recent clock corrections[6] instead of adding the most recent clock offset value to the time stamp as done in this study for reasons of simplicity.

4. Discussion

In this section we compare and discuss the results of the different timing test scenarios. As mentioned in Section 2.1.3 on the LSL framework, no smoothing of clock corrections and time stamps was applied. Hence, the results discussed here can be considered rather conservative and could still be positively influenced by applying jitter-reducing methods.

4.1. Scenario I: Sender instance timing

Timing test Scenario I showed a jitter close to the time resolution specified for the LabStreamer of 0.1 ms.[7] The results show that the PHL can accurately synchronize its audio playback with the information sent over the network using LSL. It should be noted that the lag of the PHL sender instance depends on

6 As, e.g., done in the standard MATLAB importer for .xdf files load_xdf.m: https://github.com/xdf-modules/xdf-Matlab (accessed January 14, 2022).

7 Neurobehavioral Systems. LabStreamer https://www.neurobs.com/menu_presentation/menu_hardware/labstreamer (accessed March 3, 2022).

the sampling rate-fragment size combination. A larger block size with the same sampling rate leads to a higher lag, while the jitter is expected to remain unchanged.

4.2. Scenarios II and III: EEG system timing vs. in-the-loop timing

In Scenarios II and III, a jitter of around 3 ms was measured. The increase in jitter compared to Scenario I can most likely be attributed to the EEG system, as it was not used in the first scenario. Variations of the lag from measurement to measurement occurred in both Scenarios II and III involving the EEG system. Even if not further investigated here, these findings align with previous findings when using the Smarting EEG system, which indicates that timing can vary with device, session, and software versions (Debener et al., 2015; Blum et al., 2017). Similar EEG hardware was used in Blum et al. (2017) and Hölle et al. (2022). These studies do not report any drift behavior. One potential reason for this is that in these studies, the *EEG LSL stream* was recorded on the same smartphone, which created the *EEG LSL stream*. In this study, however, the *EEG LSL stream* was recorded by an external device, i.e., the PHL. If the *EEG LSL stream* is recorded on the smartphone, no additional clock correction between PHL and EEG system is required when aligning audio and EEG.

We observed negative latencies in Scenario II. The rising edges in the *EEG LSL stream* were detected before the rising edges contained in the analog test signal. One possible explanation for this behavior is that LSL features the possibility to shift the locally created LSL time stamps by an arbitrary amount, e.g., to compensate for a known latency. If one assumes that the timestamps are generated on the smartphone, a possible explanation for the negative lags is an overcorrection of the timestamps to compensate for the latency caused by the Bluetooth connection. However, the proprietary protocol used for synchronization between Smarting amplifier and smartphone is not accessible, such that the procedure used to convert the data sent *via* Bluetooth into an LSL stream remains unknown. Hence, a more detailed investigation could not be carried out.

Scenarios II and III show very similar latency behavior over time. The two scenarios differ in the fact that Scenario II uses the reference audio signal from the LabStreamer, and Scenario III uses the reference audio signal from the PHL. In both cases, this audio signal is fed into the EEG system. The similar latency-time trend, i.e., a positive, non-monotonic drift, between Scenario II and III indicates that the EEG is causing this latency drift behavior.

The jitter obtained in Scenario II is higher than in Scenario III, presumably attributed to the higher latency fluctuations in the first approx. 200 s of the measurements. As these fluctuations did not occur in Scenarios I and III, they could be traced back to the LabStreamer's function to receive the LSL stream from the EEG system, which may differ from the PHL. One possibility might be that the LabStreamer could have applied online smoothing methods to the clock correction offset values and time stamps received from the EEG system. To record the latencies as purely as possible, the LabStreamers's settings option "Linearize Timestamps" was disabled. This option can be used to reduce the timestamps' jitter by assuming a regular sampling rate of incoming LSL streams (see Section 2.1.3). However, it is unclear if the LabStreamer still applies online smoothing methods to the clock correction offset values. Thus, the initial fluctuations of the latency could be a result of applying online smoothing using a very limited history of clock correction offset values. The exact cause for the fluctuations remains unclear; a more detailed investigation would be interesting but lies outside the scope of the current study.

4.3. Comparison with reference data recording

The comparison of the data acquired by the openMHA software with the LabRecorder has shown that data acquisition can accurately be made *via* both ways, i.e., recording with the LabRecorder as well as real-time capturing with the openMHA framework.

4.4. Suitability and comparison to state-of-the-art systems

In order to compare the temporal precision of the presented system to existing state-of-the-art technology, it is important to note that there is no standard at which point a mobile EEG system is considered sufficiently precise. Nevertheless, there are some indications that can be used for the evaluation of the system's performance. For instance, Williams et al. (2021) established a jitter threshold, i.e., the point at which jitter made an event-marking method unreliable using an ERP data set recorded using a research-grade EEG system. Their data set contained P100, N100, and P200 peaks. They determined a jitter threshold of 45 ms for the temporally wider P200 peak and a jitter threshold of 16 ms for the temporally sharper N1 peak. They concluded that larger auditory ERP peaks are more robust to jitter, while smaller peaks are more easily mitigated. While the authors point out that these values should not be regarded as absolute thresholds, they can be used as guidelines to narrow down the order of magnitude of temporal requirements for mobile EEG systems. The jitter of the system presented here is well below the jitter thresholds determined in Williams et al. (2021). Further, when comparing the results to existing studies

that report timing test data, e.g., Debener et al. (2015), Blum et al. (2017), Mirkovic et al. (2019), Williams et al. (2021), Hölle et al. (2022), the system shows a similar or better accuracy. It can therefore be assumed that the system is suitable to be used for the measurement of temporally sharper peaks, such as the N1. Additionally, in another study, the portable components used for this setup have already been used to record physiological data that enabled the extraction of ERPs, including N100 and P300 peaks, using an Oddball paradigm (Dasenbrock et al., 2021).

4.5. Limitations and challenges

It was shown that the jitter of the setup is sufficiently small and thus suitable for online applications. However, the across-session variation of the lag could have a negative influence when averaging over sessions and subjects. The respective impact of the across-session variation of the lag depends on which EEG feature is examined. Temporally wide features that can be spread over several 100 ms, such as the P300 amplitude, can still be decoded between sessions. This can be done using decoders based on 50 ms or 100 ms time bins (Debener et al., 2015; Dasenbrock et al., 2021). Nevertheless, an EEG system with higher lag predictability would facilitate its usability for future applications of the setup.

This study did not examine the extent to which the computational load of PHL might have an impact on the setup's timing accuracy. Factors such as sophisticated signal processing and the addition of more sensors could increase the PHL's CPU or network load. Thus, future studies using the PHL should always investigate the timing for their particular setup. Quick timing tests may be derived from the methods introduced here.

Due to the necessity of a PHL device, the presented setup may be harder to reproduce when compared to fully smartphone-based approaches. However, considering the PHL's specialized hearing aid hardware and openMHA signal processing software, less effort will be required to implement new EEG-based hearing aid algorithms. Regarding their form factor, both PHL and cEEGrid are not entirely suitable for everyday use. In a research context, however, the presented setup provides a useful platform to explore the potential of neuro-steered hearing devices.

4.6. Relevance for future applications

Studies such as Zink et al. (2017) or Aroudi et al. (2021) require synchronizing the incoming EEG stream with the audio stream online, as they are received over LSL, e.g., to compare the estimated envelope calculated from the EEG with the envelope of the audio. This requires on-the-fly time synchronization of incoming LSL streams, which is possible with the presented setup, as all information for synchronization is available in the block-based real-time processing in the openMHA software. Further, other approaches exist that incorporate information other than microphone signals into hearing aid processing, based on, e.g., electrooculography (EOG) and head movement sensors to estimate the user's gaze direction (Favre-Felix et al., 2018; Grimm et al., 2018). Since the presented setup uses the well-established LSL framework, other sensors could be integrated in the same manner as employed in this study. This would also allow the setup to be extended to include other sensors besides EEG to expand the setup into a multi-modal research platform (Blum et al., 2021).

5. Summary and conclusions

This study performed a comprehensive timing analysis of the research platform introduced in Dasenbrock et al. (2021), focusing on the alignment of audio and EEG data. The setup combines the mobile Smarting EEG system with a portable research hearing device, the Portable Hearing Laboratory (PHL). To perform the timing analysis, we further developed the setup to enable sending single software event markers during the onset of an acoustic stimulus. The temporal precision of the PHL when presenting acoustic stimuli and the EEG system when providing synchronized EEG data was measured using a reference device, i.e., the LabStreamer. Further, after the accuracy of the PHL's stimulus presentation and the EEG system were determined separately, the entire system was examined "in-the-loop" to quantify how the setup's timing accuracy would be in an actual EEG measurement. The data received and recorded with the presented setup were compared to the widely used standard recording program LabRecorder. All timing tests were performed for a short (15 min) and long (3 h) measurement duration. Based on the data collected in this study, we concluded the following:

- The PHL can time-accurately present acoustic stimuli and generate LSL streams over multiple hours of playback.
- The timing accuracy of the EEG system on its own can have a major influence on the overall system's timing. Checking the timing behavior of the EEG system by comparing its LSL stream against a trusted reference is a crucial step when integrating such a system. While the EEG system used in this work is sufficiently accurate within a measurement session, there are noticeable lag variations across sessions. The current setup should be enhanced with an EEG system featuring a higher temporal across-session lag stability to improve its practicability.
- The PHL is capable of presenting acoustic stimuli while simultaneously capturing the audio and EEG LSL streams with sufficient temporal accuracy over multiple hours of

playback and recording in an in-the-loop system. The temporal accuracy is sufficient to extract event-related potentials from the EEG.
- Featuring a high temporal precision and real-time signal processing capabilities, the presented setup is suitable as a platform to investigate closed-loop EEG & audio applications for future hearing aids.

Data availability statement

The data and analysis scripts are publicly available. The raw dataset can be found at https://zenodo.org/record/6857372#.YtWUwi8Rr0o. The analysis scripts as well as the openMHA configurations used in timing test Scenario III are available at https://github.com/steffendasenbrock/SynchronizationEarEEGAudioStreams.

Author contributions

SDa conceived the experiments with the help of SB, SDe, and HK. SDa carried out the data collection and analysis and wrote the manuscript with the input of SB, PM, SDe, VH, and HK. SDa, PM, and HK implemented the presented setup. All authors contributed to the article and approved the submitted version.

Funding

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2177/1 – Project ID 390895286.

Acknowledgments

We thank Lisa Straetmans and Kamil Adiloğlu for their technical support to conduct the measurements with the LabStreamer.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.904003/full#supplementary-material

References

Aroudi, A., and Doclo, S. (2019). "Cognitive-driven binaural LCMV beamformer using EEG-based auditory attention decoding," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Brighton: IEEE), 406–410. doi: 10.1109/ICASSP.2019.8683635

Aroudi, A., Fischer, E., Serman, M., Puder, H., and Doclo, S. (2021). Closed-loop cognitive-driven gain control of competing sounds using auditory attention decoding. *Algorithms* 14, 287. doi: 10.3390/a14100287

Bernarding, C., Strauss, D. J., Hannemann, R., and Corona-Strauss, F. I. (2012). "Quantification of listening effort correlates in the oscillatory EEG activity: a feasibility study," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (San Diego, CA: IEEE), 4615–4618. doi: 10.1109/EMBC.2012.6346995

Bleichner, M. G., and Debener, S. (2017). Concealed, unobtrusive ear-centered EEG acquisition: cEEGrids for transparent EEG. *Front. Hum. Neurosci.* 11, 163. doi: 10.3389/fnhum.2017.00163

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

Blum, S., Debener, S., Emkes, R., Volkening, N., Fudickar, S., and Bleichner, M. G. (2017). EEG recording and online signal processing on android: a multiapp framework for brain-computer interfaces on smartphone. *Biomed Res. Int.* 2017, 3072870. doi: 10.1155/2017/3072870

Blum, S., Hölle, D., Bleichner, M. G., and Debener, S. (2021). Pocketable labs for everyone: synchronized multi-sensor data streaming and recording on smartphones with the lab streaming layer. *Sensors* 21, 8135. doi: 10.3390/s21238135

Callaway, E., and Halliday, R. A. (1973). Evoked potential variability: effects of age, amplitude and methods of measurement. *Electroencephalogr. Clin. Neurophysiol.* 34, 125–133. doi: 10.1016/0013-4694(73)90039-4

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

Dasenbrock, S., Blum, S., Debener, S., Hohmann, V., and Kayser, H. (2021). A step towards neuro-steered hearing aids: integrated portable setup for time-synchronized acoustic stimuli presentation and EEG recording. *Curr. Direct. Biomed. Eng.* 7, 855–858. doi: 10.1515/cdbme-2021-2218

Debener, S., Emkes, R., De Vos, M., and Bleichner, M. (2015). Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear. *Sci. Rep.* 5, 16743. doi: 10.1038/srep16743

Favre-Felix, A., Graversen, C., Hietkamp, R. K., Dau, T., and Lunner, T. (2018). Improving speech intelligibility by hearing aid eye-gaze steering: conditions with

head fixated in a multitalker environment. *Trends Hear*. 22, 2331216518814388. doi: 10.1177/2331216518814388

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

Geirnaert, S., Vandecappelle, S., Alickovic, E., de Cheveigne, A., Lalor, E., Meyer, B. T., et al. (2021). Electroencephalography-based auditory attention decoding: toward neurosteered hearing devices. *IEEE Signal Process. Mag.* 38, 89–102. doi: 10.1109/MSP.2021.3075932

Grimm, G., Kayser, H., Hendrikse, M., and Hohmann, V. (2018). "A gaze-based attention model for spatially-aware hearing aids," in *Speech Communication; 13th ITG-Symposium* (Oldenburg), 1–5.

Hairston, W. D., Whitaker, K. W., Ries, A. J., Vettel, J. M., Bradford, J. C., Kerick, S. E., et al. (2014). Usability of four commercially-oriented EEG systems. *J. Neural Eng.* 11, 046018. doi: 10.1088/1741-2560/11/4/046018

Haro, S., Rao, H. M., Quatieria, T. F., and Smalt, C. J. (2022). EEG alpha and pupil diameter reflect endogenous auditory attention switching and listening effort. *Eur. J. Neurosci.* 54, 8225–8248. doi: 10.1111/ejn.15616

Hölle, D., Blum, S., Kissner, S., Debener, S., and Bleichner, M. G. (2022). Real-time audio processing of real-life soundscapes for EEG analysis: ERPs based on natural sound onsets. *Front. Neuroergon.* 3, 793061. doi: 10.3389/fnrgo.2022.793061

Intriligator, J., and Polich, J. (1995). On the relationship between EEG and ERP variability. *Int. J. Psychophysiol.* 20, 59–74. doi: 10.1016/0167-8760(95)00028-Q

Kayser, H., Herzke, T., Maanen, P., Zimmermann, M., Grimm, G., and Hohmann, V. (2022). Open community platform for hearing aid algorithm research: open master hearing aid (openmha). *SoftwareX* 17, 100953. doi: 10.1016/j.softx.2021.100953

Kidmose, P., Looney, D., and Mandic, D. P. (2012). "Auditory evoked responses from ear-EEG recordings," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (IEEE), 586–589. doi: 10.1109/EMBC.2012.6345999

Kothe, C., Medine, D., Boulay, C., Grivich, M., and Stenner, T. (2014). *Lab Streaming Layer*. Available online at: https://github.com/sccn/labstreaminglayer

Mikkelsen, K. B., Kappel, S. L., Mandic, D. P., and Kidmose, P. (2015). EEG recorded from the ear: characterizing the ear-EEG method. *Front. Neurosci.* 9, 438. doi: 10.3389/fnins.2015.00438

Mirkovic, B., Debener, S., Schmidt, J., Jaeger, M., and Neher, T. (2019). Effects of directional sound processing and listener's motivation on EEG responses to continuous noisy speech: do normal-hearing and aided hearing-impaired listeners differ? *Hear. Res.* 377, 260–270. doi: 10.1016/j.heares.2019.04.005

Obbard, C., and James, D. (2018). "Preempt_rt isn't just for lasers: the perfect match for hearing aid research!" in *Linux Audio Conference* (Edinburgh).

O'Sullivan, J., Chen, Z., Herrero, J., McKhann, G. M., Sheth, S. A., Mehta, A. D., et al. (2017). Neural decoding of attentional selection in multi-speaker environments without access to clean sources. *J. Neural Eng.* 14, 056001. doi: 10.1088/1741-2552/aa7ab4

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

Pavlovic, C., Hohmann, V., Kayser, H., Wong, L., Herzke, T., Prakash, S., et al. (2018). Open portable platform for hearing aid research. *J. Acoust. Soc. Am.* 143, 1738–1738. doi: 10.1121/1.5035670

Reis, P. M., Hebenstreit, F., Gabsteiger, F., von Tscharner, V., and Lochmann, M. (2014). Methodological aspects of EEG and body dynamics measurements during motion. *Front. Hum. Neurosci.* 8, 156. doi: 10.3389/fnhum.2014.00156

Ries, A. J., Touryan, J., Vettel, J., McDowell, K., and Hairston, W. D. (2014). A comparison of electroencephalography signals acquired from conventional and mobile systems. *J. Neurosci. Neuroeng.* 3, 10–20. doi: 10.1166/jnsne.2014.1092

Slaney, M., Lyon, R. F., Garcia, R., Kemler, B., Gnegy, C., Wilson, K., et al. (2020). Auditory measures for the next billion users. *Ear Hear.* 41, 131S–139S. doi: 10.1097/AUD.0000000000000955

Williams, N. S., McArthur, G. M., and Badcock, N. A. (2021). It's all about time: precision and accuracy of emotiv event-marking for ERP research. *PeerJ* 9, e10700. doi: 10.7717/peerj.10700

Zink, R., Proesmans, S., Bertrand, A., Van Huffel, S., and De Vos, M. (2017). Online detection of auditory attention with mobile EEG: closing the loop with neurofeedback. *bioRxiv [preprint]* 218727. doi: 10.1101/218727

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 09 January 2023
DOI 10.3389/fnins.2022.997377

OPEN ACCESS

EDITED BY
Martin Georg Bleichner,
University of Oldenburg, Germany

REVIEWED BY
Christopher Joseph Black,
Brown University, United States
Simon Lind Kappel,
Aarhus University, Denmark

*CORRESPONDENCE
Metin C. Yarici
✉ metin.yarici16@imperial.ac.uk

SPECIALTY SECTION
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

RECEIVED 18 July 2022
ACCEPTED 09 December 2022
PUBLISHED 09 January 2023

CITATION
Yarici MC, Thornton M and Mandic DP (2023) Ear-EEG sensitivity modeling for neural sources and ocular artifacts. *Front. Neurosci.* 16:997377. doi: 10.3389/fnins.2022.997377

Ear-EEG sensitivity modeling for neural sources and ocular artifacts

Metin C. Yarici*, Mike Thornton and Danilo P. Mandic

Communications and Signal Processing, Electronic and Electrical Engineering, Imperial College, London, United Kingdom

The ear-EEG has emerged as a promising candidate for real-world wearable brain monitoring. While experimental studies have validated several applications of ear-EEG, the source-sensor relationship for neural sources from across the brain surface has not yet been established. In addition, modeling of the ear-EEG sensitivity to sources of artifacts is still missing. Through volume conductor modeling, the sensitivity of various configurations of ear-EEG is established for a range of neural sources, in addition to ocular artifact sources for the blink, vertical saccade, and horizontal saccade eye movements. Results conclusively support the introduction of ear-EEG into conventional EEG paradigms for monitoring neural activity that originates from within the temporal lobes, while also revealing the extent to which ear-EEG can be used for sources further away from these regions. The use of ear-EEG in scenarios prone to ocular artifacts is also supported, through the demonstration of proportional scaling of artifacts and neural signals in various configurations of ear-EEG. The results from this study can be used to support both existing and prospective experimental ear-EEG studies and applications in the context of sensitivity to both neural sources and ocular artifacts.

KEYWORDS

ear-EEG, forward modeling, blinks, vertical saccades, horizontal saccades, EEG artifacts, neural sources

Introduction

Electroencephalography (EEG) is a brain monitoring method that utilizes non-invasive electrodes placed on the scalp surface to extract electrical neural activity. The most common use-cases of EEG are clinical; these involve the localization and characterization of seizures related to epilepsy (Noachtar and Rémi, 2009) and the objective assessment of hearing ability in infants (Schulman-Galambos and Galambos, 1979). However, since modern EEG hardware is available in a miniaturized, portable form Abiri et al. (2019) and Wolpaw et al. (2002), EEG has also attracted attention in multiple real world applications of brain monitoring, such as brain computer interfaces (BCI).

Conventionally, EEG is recorded through an array (montage) of electrodes placed across the entire scalp, termed a scalp EEG montage. As a result of the wide coverage of the human scalp achievable with a scalp EEG montage, conventional EEG offers good spatial sensitivity to a variety of neuronal activity from across the brain surface.

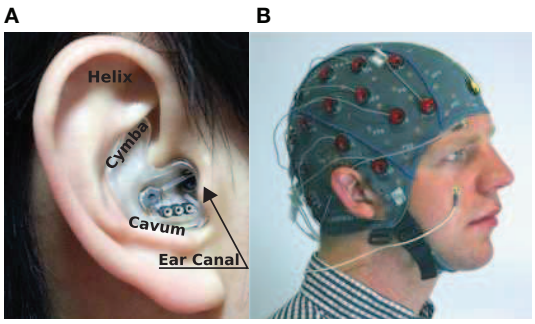


FIGURE 1
EEG sensing technologies: **(A)** In-ear-EEG device mounted on the right ear of a subject. Common locations for electrodes on the ear canal, concha (cavum and cymba), and helix are shown (Looney et al., 2012). **(B)** Standard scalp EEG cap and electrodes mounted on the head of a subject.

However, conventional scalp EEG is not suited for wearable applications as a result of the difficulty of integration of the hardware with everyday activity. Specifically, conventional scalp EEG is cumbersome, obtrusive, time-consuming to set up, difficult to use without specialist supervision, and introduces unwanted stigma for patient-populations (Casson, 2019). For these reasons, alternative, miniaturized, and wearable EEG montages which address these shortcomings hold much promise.

One such candidate is ear-EEG (Looney et al., 2011), which employs a small number of electrodes which measure EEG from the surface of the skin on the outer-ear (Mikkelsen et al., 2015). Importantly, ear worn devices are familiar, naturally discreet, unobtrusive, non-stigmatizing, and potentially easy-to-use, thus providing a convenient base for wearable health monitoring platforms (see Figure 1). Ear-EEG has been shown to be a reliable alternative to scalp EEG in several settings; sleep stage classification (Mikkelsen et al., 2017; Nakamura et al., 2017b), drowsiness onset detection (Nakamura et al., 2018), objective hearing threshold estimation (Bech Christensen et al., 2018), bio-metric authentication (Nakamura et al., 2017a), epileptic waveform detection (Zibrandtsen et al., 2017), brain-computer-interfaces (Goverdovsky et al., 2017; Yarici et al., 2021), and emotion recognition (Athavipach et al., 2019). Additionally, the susceptibility of ear-EEG to various artifacts has also been characterized experimentally for auditory neural activity detection in the presence of head, eye, and jaw movements (Kappel et al., 2017).

The source-sensor relationship is an important concept for EEG technologies. The source-sensor relationship is a characterization of the sensitivity of a sensor, or an EEG channel, to the signal sources of interest, which are typically neural current dipoles (Grech et al., 2008). In practical terms, knowledge of the source-sensor relationship not only facilitates the optimization of EEG channel configuration for a particular neural source, but can also provide rigorous, theoretical evidence for inferences drawn from weak experimental data, for example data that is collected *via* a low number of sensors and in noise-prone scenarios (Rush and Driscoll, 1968; Coburn and Moreno, 1988; Mosher et al., 1992). For wearable EEG technologies such as ear-EEG, signal-to-noise ratios (SNR) are often low in practice, which poses great challenges for the adoption of such wearable technologies in many scenarios. If the source-sensor relationship were known for wearable devices, the choice and design of the device could be optimized for particular applications.

Physics modeling of the propagation of neural potential to EEG sensors, commonly termed forward modeling, can be used to estimate the source-sensor relationship for various configurations of EEG (both montages and channels). In this paradigm, the electric potential on the surface of the scalp arising due to a current dipole source within the brain volume is estimated by applying Maxwell's equations to a structurally accurate dielectric model of the head (Sarvas, 1987), termed a volume conductor model. High resolution imaging, such as magnetic resonance imaging (MRI) scans or computerized tomography (CT) scans enable the construction of detailed volume conductor models. Unlike experimentation, forward modeling allows for the testing of a larger number of channels and sources in a time-efficient way.

Some limited forward modeling work related to ear-EEG has been reported in the literature. In an ear-EEG modeling study Kappel et al. (2019), demonstrated the feasibility of source localization using in-ear-EEG and provided highly detailed subject specific forward models that were created by scanning each subject's head anatomy. However, the source-sensor relationship for in-ear-EEG was not evaluated in detail. In a more detailed analysis, Meiser et al. employed forward modeling in order to compare the sensitivity of scalp EEG with cEEGrid–an alternative ear-EEG method which utilizes the skin surface surrounding the ear (Meiser et al., 2020). The study presented in this paper aimed to establish a detailed source-sensor relationship for various configurations of ear-EEG.

While the sensitivity to neural sources is a key determining factor in the reliability of an EEG technology, equally important is its robustness in the presence of artifacts. Despite the fact that the physics mechanisms underlying most common EEG artifacts have been long established [motion (Oliveira et al., 2016; Symeonidou et al., 2018), eye movement (Gratton, 1998; Joyce et al., 2004; Roy et al., 2014), muscle activation (Ma et al., 2012; Muthukumaraswamy, 2013; Richer et al., 2019), and electrical interference (Webster, 2009)], to date, there is no theoretical study of such artifacts in ear-EEG. In this paper, as well as exploring the source-sensor relationship for various neural sources, we aim to provide a detailed characterization of the source-sensor relationship for various ocular sources of artifacts, utilizing equivalent current dipole data for blinks, vertical saccades, and horizontal saccades [collected by Lins et al. (1993a)]. This paper aims to highlight the utility of

approaching ear-EEG equipped with theoretical knowledge of not only neural source sensitivity, but also that for sources of artifact.

Methods

Modeling was conducted in COMSOL Multiphyiscs®– a multi-physics modeling platform which enables finite element electromagnetic modeling in multiple physics domains (COMSOL, 2022). Geometric and dielectric human tissue data from the Information Technologies in Society Foundation (IT'IS; Gabriel, 1996; Iacono et al., 2015) was used to build an accurate (dielectric) model of the human head and ears to be used within COMSOL. The following section describes the structure of the model and the implementation of physics modeling within the COMSOL software.

Forward modeling

For EEG modeling, the standard volume conductor–forward modeling approach was employed (Sarvas, 1987). The same approach was also employed for ocular artifact modeling. The utilization of such an approach for ocular source modeling is motivated by the fact that the ocular sources used within this study were described as equivalent current dipoles, which were discerned through standard EEG inverse modeling procedures in Lins et al. (1993a).

In this work, potential on the surface of the skin was extracted at several EEG locations of interest. For ear-EEG, four sites on each ear were sampled; the ear canal (*XEC*), concha cavum (*XCAV*), concha cymba (*XCYM*), and helix (*XHEL*), where *X* is either *L* (Left) or *R* (Right), denoting the specific ear. The placing of the electrodes is shown in Figure 3A. For scalp-EEG, a 64-channel montage was placed on the surface of the skin according to the 10–20 BESA (2022) convention. This configuration represents a commonly used montage within EEG research and provides a reasonable density of sampling and spatial extent of coverage on the head surface.

Volume conductor model

The aim of the present study is to provide the first ear-EEG source-sensor analysis for both neural and ocular sources. As such, in addition to standard EEG forward model structures (e.g., bone, brain, skin, and muscle) a volume conductor model which includes structures of the eyes and ears is necessary. Rather than provide subject specific modeling, the current study aims to provide generalizable modeling, such that the results could be applied to the general population. When building a generalizable model, anatomical geometry which closely resembles the average anatomy of the population is desirable, as this will maximize the portion of the population for which your predictions resemble experimental measurements. However, since geometries for the ear and eyes are not used in EEG forward models or detected in routine anatomical scans of the head, such average data for the eyes and ears has not been produced. In the place of data that is generalizable, the MIDA model has been used in the current study (Iacono et al., 2015). The MIDA model describes the head anatomy of a healthy adult male, is highly detailed, and includes 153 different tissues of the head (including the vitreous humor of the eyes and the skin surface of the ears). For the present model, an adapted and simplified geometry, including seven main tissues was created (see Figure 2). The tissues included were skin, including the inner and outer dermis of the whole head, neck, and outer ear; bone, including the skull and the C1–C3 vertebrae; brain tissues including gray matter, white matter, cerebellum, and brainstem surface; internal air (respiratory tracts and mastoid air-cells within the skull); parotid glands; vitreous humor (left and right eyes); and muscle. The muscle tissue was not designated a specific geometry, rather, surrounding regions of the other tissues were endowed with muscle tissue properties.

The tissue properties used within this model are provided in Table 1 and are drawn from the IT'IS tissue frequency database (Hasgall et al., 2022). Dielectric values at an excitation frequency of 10 Hz were used for this study in order to reflect a typical EEG frequency of interest.

The COMSOL Multiphysics software requires that geometry data does not exhibit non-manifold edges and self-intersecting faces. Therefore, the MIDA data was adapted and simplified during a pre-processing procedure in order to meet these compatibility requirements. Details of the pre-processing procedure are provided in the Supplemental material, while the entirety of the final model geometry will be provided upon request.

Within the COMSOL software, forward modeling was conducted within the electric currents interface (AC/DC Module) through the following methods. An equation for current density was used to solve for electric potential throughout the model:

$$J = \sigma E + j\omega D + J_e, \tag{1}$$

where J is the total current density, σ is material conductivity, E is the electric field, j is the imaginary number, ω is the frequency of current, D is the displacement field, and J_e is the external (source) current density (COMSOL, 2022).

The following equation of continuity was imposed across tissue boundaries:

$$n_2(J_1 - J_2) = 0, \tag{2}$$

where n is a unit vector that is normal to (and directed away from) the boundary and J is the electric current density. The indices 1, 2 describe the regions of space either side of the

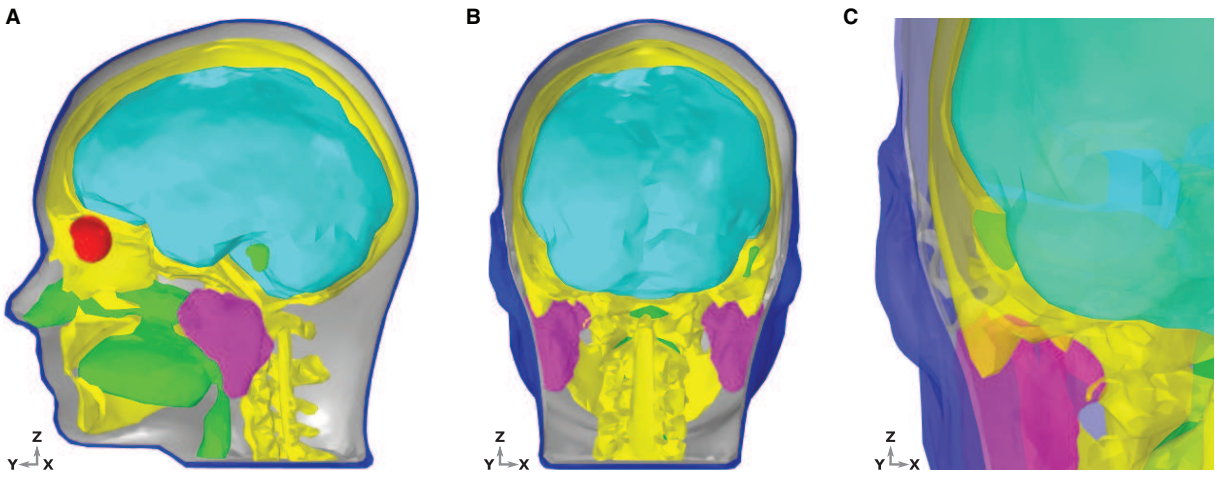


FIGURE 2
Dielectric distribution within the head and ear model. Different tissue groups are discernible by color; brain-cyan, bone-yellow, muscle-gray, air-green, vitreous humor-red, and skin-blue. **(A)** Sagittal section of the head showing the interior of the model. The skin and bones are hidden to the left of the mid line sagittal plane, such that the brain, eyes, glands, and internal air are clearly visible. **(B)** Coronal section of the head with the skin and bone hidden to the rear of the mid line coronal plane. **(C)** Coronal section of the head, zoomed in to the left ear region, viewed from a posterior position. The skin and bones are hidden to the rear of the mid line coronal plane. Tissues are transparent, to enable a detailed view of the entire ear and surrounding structures.

TABLE 1 Dielectric properties of the tissue in the model taken from the IT'IS tissue frequency database (Hasgall et al., 2022).

	Tissue						
	Bone	Skin (dry)	Brain (gray matter)	Muscle	Parotid glands	Eyes (vitreous humor)	Air
$\sigma(S/m)$	2×10^{-2}	2×10^{-2}	2.9×10^{-2}	2.2×10^{-1}	6.7×10^{-1}	1.5	0
ϵ_r	5.5×10^{-4}	1.1×10^{3}	4×10^{7}	2.6×10^{7}	9.4×10^{1}	9.9×10^{1}	1

Values for conductivity (σ) and relative permittivity (ϵ_r) at an excitation frequency of 10 Hz were used in the model.

boundary. The head model was placed at the center of a large ($r = 10$ m) sphere with the dielectric properties of air (conductivity $\sigma = 0$ μS/cm, relative permittivity $\epsilon = 1$). The surface of the sphere acted as the ground in the AC model. See Martinek et al. (2008), Pelot et al. (2018), and Seibt et al. (2019) for examples of volume conductor modeling within COMSOL.

Neural sources

The goal of the presented modeling was to evaluate the ear-EEG source-sensor relationship for a variety of realistic neural sources. The source space was restricted to the surface of the brain and comprised 990 homogeneously distributed source locations (Figure 3C). In this way, an exploration of the source-sensor relationship for a range of realistically located neural sources was achieved. Each location was occupied by a single source, enabling a fine-grained mapping of the brain surface. Each source was orientated at perpendicularly to the surface of the brain. Sources were modeled as point current dipoles; the location, orientation, and magnitude of which were specified (Sarvas, 1987).

Sensitivity maps

The source sensor relationship between a given EEG montage and the neural current sources is analyzed through the following method: for each of the 990 neural sources, the potential difference between all possible pairings of two electrodes (i.e., all possible bipolar channels) was determined. Of all these potential differences, the greatest unsigned potential difference is defined as the sensitivity of the EEG montage to the neural source in question. This method effectively identifies the best-case signal magnitude that can be measured by the EEG montage in question, for all neural current sources modeled. The neural sources were modeled as point current dipoles oscillating at 10 Hz. The entirety of the ear skin-surface is available to ear-EEG devices, meaning that EEG can be detected from anywhere on this surface through the use of multiple electrodes mounted on a single device (see the ear-EEG electrode locations in Figures 3A, B). The sensitivity maps in Figure 5 display the sensitivities for each of the 990 modeled brain surface sources. The analysis was conducted for both unilateral and bilateral ear-EEG and scalp-EEG (64 channel 10–20 montage). Sensitivity data were transferred to visual representations of the sensitivity over the surface of the brain, or sensitivity maps, in

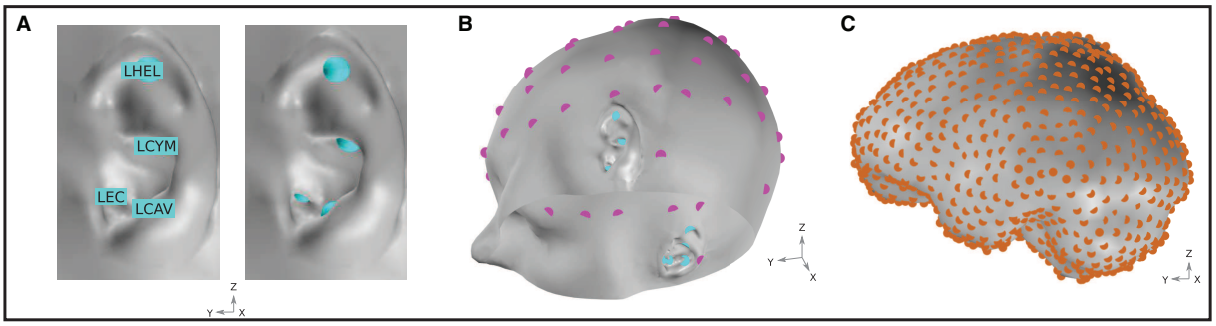


FIGURE 3
(A) Ear-EEG electrode locations and naming convention. Electrodes are shown for the left ear. An equivalent array is positioned on the right ear. **(B)** EEG electrode locations on the skin surface of the model. Magenta and cyan markers, respectively, indicate the location of electrodes in a 64-electrode scalp EEG montage and a left and right ear-EEG montage. Sixty-four scalp EEG electrodes are placed according to the 10–20 BESA (2022) convention. **(C)** Neural source space on the course-grained surface of the brain. Source locations are homogeneously distributed with a density of $1.5\,\mathrm{cm}^{-2}$ and are highlighted in orange.

MatLab®, using the source location and brain mesh data in the function `trisurf`. In order to enable clear visualization of the exponentially dynamic sensitivity map, sensitivities are plot in decibels (dB) relative to an arbitrary value, calculated through the equation $10log_{10}(V/V_{reference})$, where V is the sensitivity value of interest and $V_{reference}$ is the arbitrarily set reference value. In the chosen logarithmic scale, 10-fold differences in sensitivity are equal to $\pm$ 10 dB.

In addition to the sensitivity maps described above, relative sensitivity maps were created. In this analysis, the ear-EEG montage sensitivities were divided by the scalp EEG montage sensitivities, in order to enable calculation of the change in signal amplitude associated with the use of a particular ear-EEG montage over the 64-channel scalp-EEG montage. Once again, for each source, the scalp EEG and ear-EEG sensitivities are extracted from the optimal differential pair of electrodes within their respective montages. In this way, the expected increase/decrease in signal amplitude (signal gain/loss) associated with the use of both unilateral and bilateral ear-EEG over scalp EEG could be fairly examined. Meiser et al. introduced this method for the analysis of the cEEGrid source-sensor relationship in Meiser et al. (2020). The presently reported values are transformed into a logarithmic scale *via* the equation $10log_{10}(V_{ear}/V_{scalp})$, where V_{ear} and V_{scalp}, respectively, are the optimal sensitivity values for the ear and scalp montages.

Ocular artifact modeling

Artifacts arising due to blinks and eye movements can be explained in terms of the corneo-retinal dipole (CRD) field. This dipole field arises due to natural charge separation between the cornea and the retina. During eye movements, the CRD rotates around the center of the eyeball, resulting in a dipole current. During eye blinks, the conductive surface of the inner eyelid sweeps over the cornea, leading to current discharge, which can also be modeled as a current dipole. In Lins et al. (1993b), the authors performed dipole fits to electrooculographic data. They found that two-dipole fits (one dipole per eye) explained the data very well (explaining up to around 98% of the total variance). For each type of ocular artifact (vertical saccades, horizontal saccades, and blinks), the fitted dipoles shared approximately the same locations. For blinks, the dipoles were approximately aligned in the anterior-posterior direction. For vertical (horizontal) saccades, the dipoles were approximately aligned in the superior-inferior (lateral) direction. We therefore modeled each type of artifact using point current dipoles aligned with the directions reported by Lins et al. (1993a). Figure 4 displays a graphical representation of the dipoles, while numerical representations of the dipole vectors are provided in the Supplemental material.

For modeling ocular artifacts, a current dipole of fixed amplitude was used to model all three artifact types (blinks, vertical saccades, and horizontal saccades). The dipole amplitude was calibrated so that the resulting scalp topographies reflected the typical waveform amplitudes measured experimentally [for example, see Lins et al. (1993a)]. However, the values reported in this study are normalized and are therefore independent of the chosen dipole amplitude (the same results could be achieved with any dipole amplitude).

In our simulations, we neglected the rider artifact, which is a transient onset effect which occurs at the start of a vertical or horizontal saccade. Similar to blink artifacts, the rider artifact occurs because the eyelid lags behind the motion of the artifact, discharging slightly. In fact, the two blink dipoles can be used to explain the rider artifact.

For the purpose of validation of the ocular artifact modeling, simulations of scalp EEG sensitivities provided by the presented model were compared to measured data in Lins et al. (1993b);

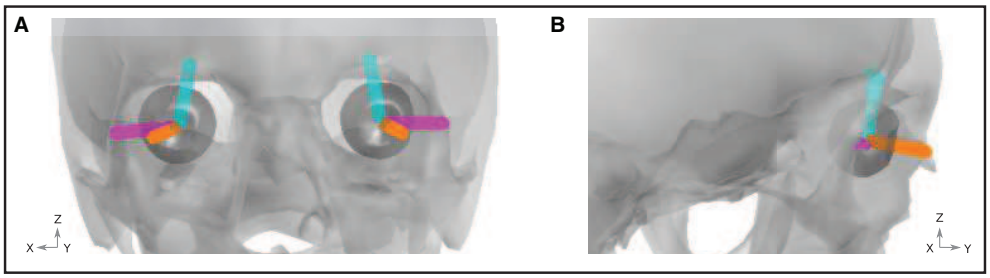


FIGURE 4
Ocular artifact dipoles. Orientation of the blink (orange), vertical (cyan), and horizontal (magenta) dipoles are shown within the geometry of the skull (gray) and eyes (dark gray). **(A)** Front view. **(B)** View from the right hand side of the head.

both sets of data are provided in Supplementary Table 3. The sensitivity of a number of scalp EEG channels ranging from frontal, central, occipital, and temporal sites were compared, where the sensitivities were expressed as percentages relative to values from reference channels; VEOG (for the blink and vertical saccade artifacts) and HEOG (for the horizontal saccade artifacts). Correspondence between the measurements in Lins et al. (1993b) and the presented simulations was calculated in terms of the mean error between sensitivity values. Good agreement was found between the measured and simulated potentials for all three of the investigated ocular artifacts, with a mean error of 3 % across all three artifacts.

Results

Ear-EEG sensitivity to neural sources

Sensitivity maps

Figures 5A, B displays the sensitivity map for a left ear unilateral ear-EEG montage (displaying the characteristic sensitivity profile for a single earpiece). As previously described, the sensitivities displayed for each individual dipole are extracted from the optimal differential pair of electrodes within the montage (for that dipole). This analysis enables examination of the full capability of a montage which is achievable through the optimal electrode pairing. The highest sensitivities of the unilateral montage were exclusively observed in the ipsi-lateral inferior and middle temporal lobe. Decreases in sensitivity were observed for regions surrounding the ipsi-lateral temporal lobe, with the lowest sensitivities observed for sources furthest away from the ipsi-lateral ear; in frontal, central, and posterior, and contra-lateral locations. For the bilateral montage, high sensitivities are observed across large portions of the left and right temporal lobe and even some surrounding regions, while the lowest sensitivities were observed for frontal, central, and posterior regions close to the mid-line sagittal plane.

The relative sensitivity of the unilateral montage is displayed in Figures 5E, F. As with the sensitivity maps described above, relative sensitivity values are calculated using the optimal electrode pairing within each montage. On the inferior ipsi-lateral temporal lobe, for a small collection of sources (2% of the total) there is moderate signal gain, with a median value of 2 dB (25/75th percentile: 1/4 dB). For the majority of the remaining sources, there is a severe signal loss. The median relative sensitivity for all sources for the unilateral ear-EEG montage is −17 dB (25/75th percentile: −20/−4 dB). For the bilateral montage, the regions of severe signal loss are reduced relative to the unilateral montage. The median relative sensitivity was found to be −10 dB (25/75th percentile: −15/−4 dB). A small portion (5%) of sources on both temporal lobes were detected with a signal gain, with the median of 2 dB (25/75th percentile: 1/3 dB).

Channel sensitivity analysis

In Figure 6, the sensitivities of ear-EEG channels are assessed individually. Channels were created between all possible pairings of electrodes within each ear-EEG montage (unilateral and bilateral). First, the prevalence of the channels is evaluated, where the prevalence is equal to the percentage of brain sources for which the channel in question recorded the highest sensitivity. The prevalence can be viewed to indicate the relative utility of a channel within its montage. This approach was introduced in Meiser et al. (2020), for channels within a unilateral cEEGrid montage. In the prevalence analysis, only sources for which at least one channel from within the montage recorded a sensitivity above a certain threshold were included in the analysis. For all montages, the threshold was set to 10% of the highest sensitivity from the scalp montage, therefore sources for which large reductions in amplitude were observed did not feature in the analysis [see Meiser et al. (2020) for a similar methodology]. The use of thresholding was motivated by the fact that for sources which are poorly detected, knowledge of the channel that recorded the highest sensitivity is not informative.

Figure 6A displays the channel prevalence in the form of a heat map for both unilateral and bilateral montages (indicated

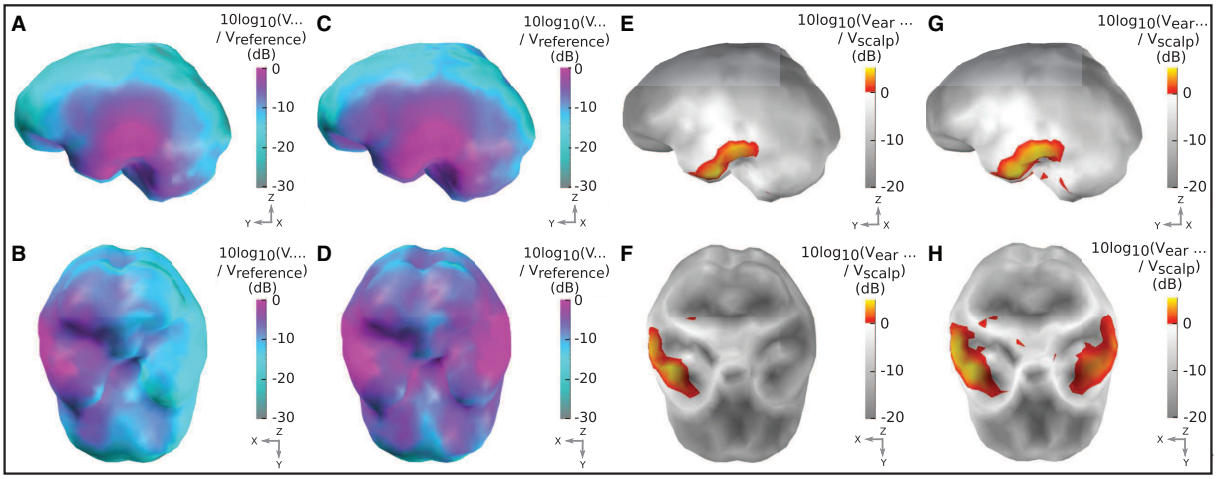


FIGURE 5
Sensitivity maps for ear-EEG: **(A, B)** Sensitivity map for a left ear unilateral ear-EEG montage. **(C, D)** Sensitivity map for a bilateral ear-EEG montage. **(E, F)** Relative sensitivity map for a left ear unilateral montage and a 64-channel scalp EEG montage. **(G, H)** Relative sensitivity map for a bilateral ear-EEG montage and a 64-channel scalp EEG montage. **(A–D)** High and low sensitivities, respectively, are represented by magenta and cyan shading. **(E–H)** Severe and moderate signal losses are displayed in gray and white, respectively. Signal gains are displayed in red and yellow. **(A, C, E, G)** Left brain surface. **(B, D, F, H)** Inferior surface of the brain.

by a dotted pattern and color shading, respectively). Within the left unilateral montage (bottom left of the heat map, separated by a dashed line) the helix to ear canal channel recorded the maximum sensitivity for 56% of selected sources. Helix to cavum, and cymba to ear canal were the next most prevalent (30 and 10%, respectively), followed by cavum to ear canal (3%). Helix to cymba and cymba to cavum failed to record the highest sensitivity for a source. Within the right ear unilateral montage (top right of the heat map, separated by a dashed line) the helix to cavum was the most prevalent (62%), with the helix to ear canal recording the highest sensitivity for the remaining sources (38%). Within the bilateral montage, which included all available electrodes on both the left and right ears, the left helix to right helix channel recorded the most maximum sensitivities (36%) for a single channel, while the bi-ear helix to ear canal channels (left helix - right ear canal/right helix to left ear canal) recorded the maximum sensitivity for 20% of sources each (Figure 6A). The left helix to right cymba and right helix to left cymba were the next most prevalent, recording 10 and 9% of maximum sensitivities, respectively. Figure 6B displays the complete set of channels which recorded the maximum sensitivity at least once. Dominance of the bilateral channels is clearly observed.

To supplement the channel prevalence analysis, the average sensitivity for each ear-EEG channel was also calculated (Figure 6C). As with the channel prevalence analysis, average sensitivities were based only on sources which satisfied the threshold condition described above. Values of average sensitivity are provided in normalized units (n.u.)–whereby unilateral and bilateral ear-EEG were normalized with respect to the same value: the maximum average ear-EEG sensitivity from within the unilateral and bilateral montages. For the average sensitivities, a linear scale was sufficient to reveal meaningful trends, i.e., the values reported are calculated *via* the equation V_{ave}/V_{max}, where V_{ave} is the average sensitivity of the channel in question, and V_{max} is the maximum average sensitivity. The helix to helix channel recorded the highest average signal amplitude (1 n.u.), however the majority of bi-ear channels recorded similar average amplitudes (>0.6 n.u.). The single ear channels exhibited lower average signal amplitude (<0.3 n.u.); the lowest average amplitude was recorded by the left and right single ear cavum to ear canal channels (<0.1 n.u.).

In order to further characterize ear-EEG, for a selection of ear- and scalp EEG channels, channel sensitivities were plotted against respective inter-electrode distance (Figure 6D). In order to comparison between scalp EEG and ear-EEG, the sensitivity of each channel was measured as the number of sources for which the channel sensitivity exceeded the previously described threshold. Ear and scalp channels were normalized with respect to the same value - the maximum sensitivity from within the ear-EEG and scalp EEG montages. For the scalp EEG channels, a linked mastoid referencing system was used, while for ear-EEG, left and right ear channels were referenced to the ipsi-lateral helix, and bi-ear channels to the contra-lateral helix. The inter-electrode distance for the linked mastoid referenced scalp EEG channels was calculated as the average of the distance of the primary electrode from both mastoid electrodes.

Both the sensitivity and inter-electrode distance of single ear channels are lower than those of the bi-ear and scalp EEG channels, leading to the ratio of mean channel sensitivity for left ear, right ear, and bi-ear-EEG relative to the mean channel

sensitivity of scalp EEG, respectively, of 0.3, 0.3, and 0.9. A linear trend with equation of best fit, $y = 0.1x + 1$, ($R^2 = 0.56$) was found for the data displayed in the plot, where y = sensitivity, x = inter-electrode distance, and R^2 is the goodness-of-fit measure for the linear trend.

Ocular artifact modeling

The sensitivity of different configurations of ear-EEG in the presence of three common types of ocular artifact, blinking, vertical saccade, and horizontal saccade, was investigated. The sensitivity of ear-EEG to ocular artifacts was bench-marked against that of scalp EEG in Figures 7A, C, E. Specifically, sensitivities for two ear-EEG channels; single-ear left helix to left ear canal (LEH-LEC) and bi-ear left helix to right helix (LHEL-RHEL) are displayed alongside sensitivities of multiple scalp EEG channels that were referenced to the linked mastoids. Since the scalp EEG channel sensitivities are approximately symmetric about the midline sagittal plane, the inclusion of only a single hemisphere's (left) EEG channels was sufficient to capture the general variations in scalp EEG sensitivities to ocular artifacts. As a result of the large range (multiple orders of magnitude) of channel sensitivities, values in dB were calculated *via* the equation $[10log_{10}(V/V_{reference})]$, where V is the channel sensitivity of interest, and $V_{reference}$ is equal to an arbitrarily set reference value. For all three artifacts, sensitivity values are reported with respect to the same reference value to enable between-artifact comparison.

For the blink artifact (Figure 7A), the maximum scalp EEG sensitivity was recorded by the FP1 channel and the lowest by the Iz channel (33 dB difference). Regarding ear-EEG, the blink artifact resulted in a larger potential difference in the LHEL-LEC channel relative to the LHEL-RHEL channel, with a difference between the sensitivities of 3 dB. Relative to scalp EEG, the LHEL-LEC channel was most similar (<1 dB difference) to scalp channels with a lateral positioning (P7, TP7), while the LHEL-RHEL channel was most similar (<1 dB difference) to the posterior scalp channel O1. The LHEL-LEC and LHEL-RHEL sensitivities were among the least sensitive out of the selection of scalp and ear-EEG channels analyzed (7th and 2nd least sensitive out of 37 channels, respectively). The resultant potential topography on the surface of the head with overlaid EEG channel topography is shown in Figure 7B, while a magnified view of the potential on the ears is provided in Figure 7G.

For the vertical saccade artifact, the maximum scalp EEG sensitivity was recorded by the FP1 channel and the lowest by the Iz channel; (26 dB difference). The LHEL-LEC channel was most similar to lateral and posterior scalp EEG channels, P7 and POz (<26 dB difference), while the LHEL-RHEL channel was most similar to the posterior inferior scalp EEG channel, Iz (<5 dB difference). The LHEL-LEC channel was 14 dB more sensitive relative to the LHEL-RHEL channel. The LHEL-LEC and LHEL-RHEL sensitivities were among the least sensitive out of the selection of scalp and ear-EEG channels analyzed (6th and 1st least sensitive, respectively). The topography of the vertical saccade potential over the surface of the scalp and ear, respectively, are shown in Figures 7D, H.

For the horizontal saccade artifact, the maximum scalp EEG sensitivity was recorded by the AF7 channel and the lowest by the central and parietal scalp EEG channels along the midline sagittal plane, Cz and CPz (<2 dB difference), while the LHEL-RHEL channel was most similar to frontal-lateral and central-lateral channels, F5 and FC5 (<2 dB difference). The LHEL-LEC channel was 15 dB less sensitive relative to the LHEL-RHEL channel. The LHEL-LEC sensitivity was among the least sensitive out of the selection of scalp and ear-EEG channels analyzed, while the LHEL-RHEL sensitivity was among the most sensitive (4th least sensitive and 6th most sensitive, respectively). The topography of the horizontal saccade potential over the surface of the scalp and ear, respectively, are shown in Figures 7F, I. The mean, maximum, and minimum sensitivity for the blink, vertical saccade, and horizontal saccade artifact were also calculated for each montage and are presented in Table 2. For each artifact, Figures 7J–L display the normalized sensitivities for the various ear-EEG channels, where for each artifact, the channel sensitivities were normalized *via* the equation $V_{channel}/V_{max}$, where $V_{channel}$ is the sensitivity of the channel in question, and V_{max} is the maximum sensitivity from all evaluated channels.

Discussion

Sensitivity to neural sources

The sensitivity of both unilateral (single ear) and bilateral (bi-ear) montages to neural sources across the entire brain surface was examined. While unilateral montages are confined to measuring potential differences over the small region of the ear, bilateral montages enable measurement between the left and right ears. In this way, the bilateral montage increases the inter-electrode distance, and therefore the potential difference, as a result of the physical laws governing EEG. Indeed, in Kappel et al. (2019), such differences are clearly observed between exemplar single channel lead fields for unilateral and bilateral ear-EEG. However, the present results for the montage sensitivities, which show the optimized sensitivity over a more diverse range of sources over the brain surface, reveal that several key, large scale variations in sensitivity for montages are similar in the unilateral and bilateral cases. For example, in (i) temporal lobe regions of highest sensitivity in close proximity to the ear electrodes and (ii) the regions of lowest sensitivity in proximity to the midline sagittal plane, the sensitivity profiles are similar. However, in between these regions, benefits of larger

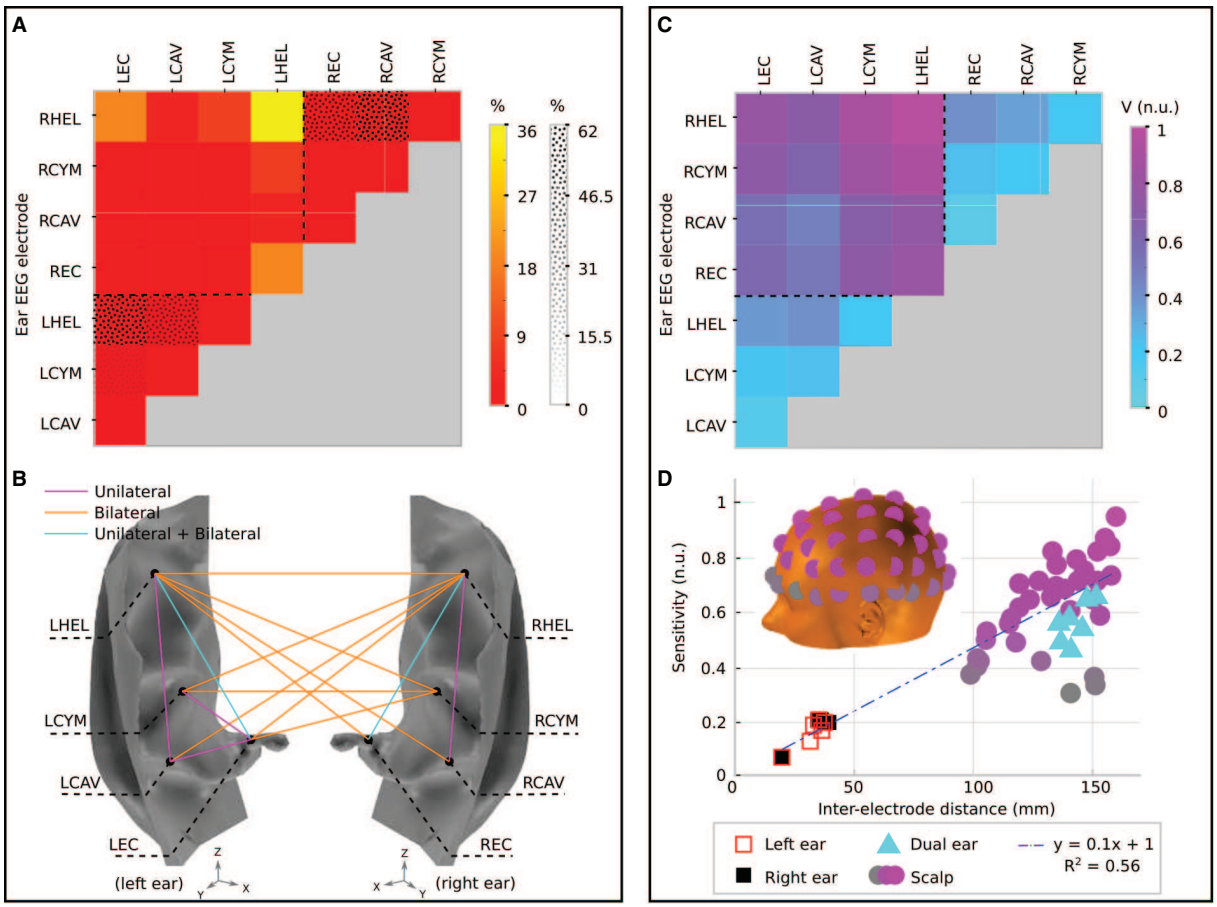


FIGURE 6
Channel prevalence and average sensitivity analysis for unilateral and bilateral montages of ear-EEG. **(A)** Channel prevalence for unilateral and bilateral montages of ear-EEG. Bilateral prevalence is shown in a color scale, while unilateral prevalence for the left and right ears is shown a in texture scale. Dashed lines separate single ear and bi-ear values on the heat-map. **(B)** Channel prevalence for unilateral and bilateral ear-EEG montages visualized on the surface of the ear. For channels that recorded the highest sensitivity within a montage for at least one source (>0%), a colored line connects the channel's electrodes. Magenta lines connect channels that were prevalent within the unilateral (left or right) montage analysis; orange lines connect prevalent channels in the bilateral montage analysis; and cyan lines connect channels which appeared prevalent in both unilateral and bilateral montage analyzes. **(C)** Normalized average sensitivity for unilateral and bilateral montages. Dashed lines separate single ear and bi-ear values on the heat-map. Unilateral and bilateral ear-EEG signals were normalized with respect to the same value. **(D)** Relationship between normalized sensitivity and inter-electrode distance for both ear-EEG (unilateral and bilateral) and scalp EEG. Both scalp EEG and ear-EEG were normalized with respect to the same value. A linear fit of the data is shown in a blue dashed line. Scalp EEG channel locations are indicated in the inset and colored corresponding to their sensitivity values. All of the channels that are displayed are referenced to the average of the sensitivities of the mastoid electrodes. The location of the left mastoid electrode is indicated by a black circle on the inset, while the right mastoid electrode is hidden from view (on the equivalent position on the right side of the head).

inter-electrode distance were clearly observed. Such differences between the unilateral and bilateral montages were observed for both the regular sensitivity maps (ear-EEG sensitivity) and the relative sensitivity maps (ear-EEG sensitivity bench marked against scalp EEG sensitivity).

Since EEG is conventionality recorded through scalp EEG, it is useful to compare the amplitude of the ear-EEG signal to that of scalp EEG. Therefore, relative sensitivity maps were also created. Ear-EEG produced an increase in signal amplitude in small regions in the temporal lobe, while adjacent regions mostly exhibited a moderate decrease in signal amplitude. In regions furthest away from the ear-EEG electrodes, the ear-EEG amplitude was shown to be considerably smaller than that of scalp EEG. The results for sources in and adjacent to the temporal lobe suggests that ear-EEG can be expected to record EEG amplitudes similar to those seen in scalp EEG in these regions. Since temporal lobe neural activity is known to correspond to important auditory and visuo-auditory processing, among other functionality, the use of ear-EEG in applications such as enhanced, "smart" hearing aids is strongly supported by these results. Indeed, reliable hearing threshold estimation on subjects with normal hearing and sensorineural

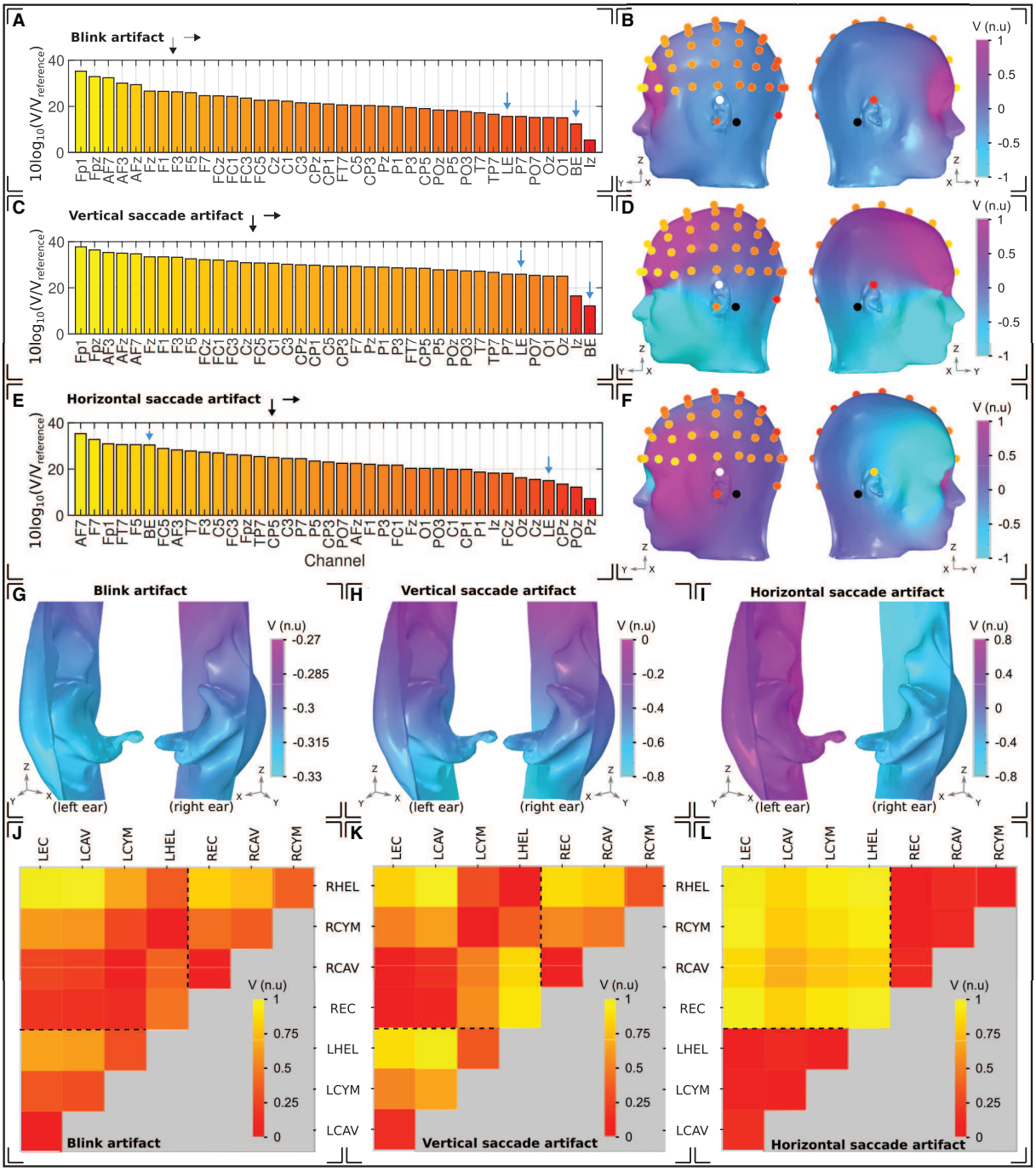


FIGURE 7
Ocular artifacts. **(A, C, E)** Sensitivities of the left hemisphere scalp EEG channels in addition to a unilateral left ear (LE) and bilateral (BE) ear-EEG channel, for **(A)** blinks, **(C)** vertical saccades, and **(E)** horizontal saccades. **(B, D, F)** Head surface potential topography and EEG sensitivities in normalized units (n.u.) arising due to **(B)** blinks, **(D)** vertical saccades, and **(F)** horizontal saccades. Scales are normalized uniformly across each of the head surface topographies such that an inter-artifact comparison is possible. White (black) circles indicate ear (scalp) reference electrodes. **(G–I)** Left and right ear surface potential topography arising due to **(G)** blinks, **(H)** vertical saccades, and **(I)** horizontal saccades. Potential scales are shared with the scalp topographies, however the color scale has been changed to enable clear visualization of the potential topography over each ear surface. **(J–L)** Sensitivity in normalized units (n.u.) of ear-EEG channels arising due to **(J)** blinks, **(K)** vertical saccades, and **(L)** horizontal saccades. The potential scale has been normalized for each individual plot; potential scales are not shared between artifacts.

hearing loss was demonstrated in two studies (Christensen et al., 2017; Bech Christensen et al., 2018), where comparable variance in scalp and ear-based estimations was found.

Sensitivity analysis of the various possible channels within a unilateral montage showed that the channels which maximize the space available on the ear surface, while also utilizing the helix position/electrode (at closer proximity to the brain), recorded the highest amplitude for the majority of sources. These channels were: the helix to ipsi-lateral ear canal and helix to ipsi-lateral cavum channels. These results are in line with experimental data in Kappel et al. (2016), where optimum reference configurations for the auditory steady-state response (ASSR) were investigated, and theoretical predictions in Meiser et al. (2020), where forward models of cEEGrid showed that channels maximizing the vertical distance between electrodes possessed the most favorable sensitivity profiles. For the bilateral montage, the same trend was observed, with the helix to contra-lateral ear canal channels and helix to contra-lateral helix channels producing the highest sensitivities. As such, in the scenario where a reduction of the ear-EEG montage size is desirable, the use of the helix and ear canal electrodes could be prioritized. However, despite the dominance of the helix and ear canal electrodes, in both the unilateral and bilateral montages, multiple electrode locations contributed to the highest sensitivity for at least one source, indicating that, while the area available on the ear surface is small, varied channel geometry within the small area is beneficial. In other words, several configurations of ear-EEG could be of use often in practice. Such benefits of varied channel geometry have been experimentally demonstrated in Kappel and Kidmose (2017), where high-density ear-EEG earpieces, with electrodes covering a large area of the ear surface, were tested in the presence of visual and auditory responses. The authors showed that for each different response, several locations across the ear could be used with comparable performance, while between responses, the optimal locations for EEG detection varied. Benefits of a high density ear-EEG array were also demonstrated in Kappel et al. (2019), where subject specific volume conductor model predictions and experimental high density ear-EEG data were shown to be in good agreement.

To provide further insight, the average sensitivity of the various possible channels within a bilateral and unilateral montage was also calculated. Calculations further supported the use of multi-electrode ear-EEG array, through comparable average sensitivities for geometrically similar channels. The expected signal amplitude of various ear-EEG and scalp EEG channels were then examined in relation to their inter-electrode distance. In general, for both scalp and ear-EEG, the expected signal amplitude was linearly proportional to inter-electrode distance. The revealed relationship between expected signal amplitude and inter-electrode distance could be used during the design process for various wearable/Hearable devices which utilize the ear and scalp surfaces. Further, it shows that even for ear-based EEG devices, amplitude will scale with area covered on the ear.

Sensitivity to ocular artifacts

For prospective general purpose EEGs, the sensitivity to ocular artifacts is an important factor to consider. During most everyday activities, humans frequently blink and perform visual scans which involve horizontal and vertical saccades, with each instance of eye movement presenting a different artifact in the EEG signal. The sensitivity to ocular artifacts for ear-EEG and scalp EEG was demonstrated through both single channel sensitivity calculations and topographical plots of potential over the surface of the head. Ear-EEG sensitivities were calculated for a characteristic channel from both a unilateral montage (single ear) and a bilateral montage (bi-ear), while multiple linked mastoid reference scalp-EEG sensitivities were also calculated. Generally, the sensitivity to ocular artifacts for the single ear channel relative to the linked mastoid scalp EEG channels was observed to be low, evidenced through channel sensitivities of ear-EEG matching those of scalp EEG channels which are among the least severely affected by ocular artifacts. There was further evidence of low ear-EEG sensitivity in the topographical plots, where, relative to the scalp surface, the ear surfaces exhibited small ranges of potential. Experimentally measured EEG SNR deterioration caused by ocular artifacts was investigated in Kappel et al. (2017). For blink artifacts, deterioration in SNR was observed in multiple scalp electrodes, while for ear-EEG electrodes, no deterioration was detected. These results are in good agreement with the findings in this paper, as well as previous scalp EEG studies (Lins et al., 1993b; Gratton, 1998; Joyce et al., 2004; Roy et al., 2014). For vertical and horizontal eye movements, unexpected SNR deterioration patterns in scalp-EEG were observed. While vertical and horizontal saccades were performed by subjects at a rate of once every 4 s (0.25 Hz), the only deterioration in EEG for the scalp electrodes was observed in the theta- to gamma-EEG range (4–30 Hz). Since EMG activity is most prominent in higher frequency ranges, the authors attributed the unexpected patterns of EEG deterioration to inadvertent muscle contractions during the measurements, as opposed to the investigated ocular artifacts. Although deterioration in the delta- to gamma-EEG range (0–30 Hz) was observed for ear-EEG, further measurements which support these findings are required. The experimental difficulties highlighted in Kappel et al. (2017) demonstrate the importance of biophysics modeling approaches, which enable the investigation of electrophysiological sources in isolation.

For the bi-ear ear-EEG, the sensitivity was also low for the blink and vertical saccade - evidenced through

TABLE 2 Sensitivity values for scalp EEG and ear-EEG montages in response to blink, vertical saccade, and horizontal saccade artifacts.

Sensitivities to ocular artifacts: $10log_{10}(V/V_{reference})$ (dB)			
	Ocular artifact		
EEG montage	Blink	Vertical saccade	Horizontal saccade
Scalp (linked mastoid reference)	26 (35, 5)	31 (38, 17)	27 (35, 7)
Left and right unilateral ear-EEG	13 (16, 1)	24 (27, 15)	20 (22, 14)
Bilateral ear-EEG	13 (17, 2)	23 (26, 10)	30 (31, 29)

For each montage, mean values are reported alongside maximum and minimum values, respectively, in parentheses. Values are reported in dB, *via* the transform $10log_{10}(V/V_{reference})$, where V is equal to the channel sensitivity, and $V_{reference}$ is a reference value common to all artifacts and EEG montages/channels.

similarity to low sensitivity scalp EEG channels and minimally varying topography over the ear skin-surfaces. However, an exception occurred during the horizontal saccade, when sensitivities equal to those of the worst affected frontal-lateral scalp EEG channels were observed for the bilateral channel. Such results are explained by the topography for the horizontal artifact (Figures 7F, I), which reveals a large potential difference between the two ear surfaces. Measurements of such bi-ear ear-EEG data that could be compared to the presented simulations have not yet been reported.

The variation in ocular artifact amplitude between the various possible ear-EEG channels was also examined. The map-plots of sensitivity show expected trends, where channels with larger inter-electrode distance have generally larger amplitudes. However, there are also variations in sensitivity which arise due to variations in channel orientation relative to the dipole field. For example, for the vertical saccade artifact, there is a 14 dB increase in sensitivity for the LHEL-LEC channel relative to the LHEL-RHEL channel, despite a much smaller inter-electrode distance (Figure 7C). Such characteristic variations within the ear-EEG montages could be used in the detection of artifacts, and highlight another benefit of utilizing a multi-electrode ear-EEG array, as opposed to single channels. Ear-EEG data that could be compared to the presented channel sensitivity analysis of ocular artifacts have not yet been reported.

Suggestions for ear-EEG

For the first time, a systematic and detailed analysis of the ear-EEG source-sensor relationship was provided for a wide variety of neural sources from realistic locations, while various configurations of ear-EEG were considered. In addition, novel ear-EEG source-sensor relationships for vertical saccade, horizontal saccade and blink-related ocular artifacts have been established. With regard to both EEG detection and sensitivity to ocular artifacts, such source-sensor mapping, while serving to provide novel insight into the ear-EEG sensitivity profile, could also be used to support experimental measurements from ear-EEG in existing literature or in prospective studies. Additionally, the methods employed within this study can be adopted with reasonable ease by researchers for the purpose of conducting new ear-EEG simulations.

Simulations of neural source sensitivity have conclusively supported the use of both unilateral and bilateral ear-EEG montages for the detection of neural activity originating from within the temporal lobe. In both the unilateral and bilateral cases, the ear-EEG was estimated to record higher or similar amplitudes to conventional scalp EEG within these regions. This suggests that existing protocols for EEG detection could be used with ear-EEG, without the need for considerable changes to the protocol, and with equally likely success. In fact, as a result of the wearability of ear-EEG, existing auditory EEG protocols could feasibly be enhanced to include more novel real world recording scenarios, as demonstrated in a "smart helmet with ASSR" study (Von Rosenberg et al., 2016), where auditory brain responses were recorded from a subject while riding a bike. Since moderate decreases in amplitude were also observed for ear-EEG in brain regions adjacent to the temporal lobe, covering a variety of neural function, there is also support for similar use of ear-EEG in a wider range of applications.

Despite the models predictions of low amplitude ear-EEG for sources located in areas furthest away from the temporal lobes, e.g., the posterior regions of the brain, there is experimental support for ear-EEG detection of neural activity from such regions, such as the visual cortex in the occipital lobe (e.g., Kidmose et al., 2013; Goverdovsky et al., 2017). In these experimental studies, the successful detection of visual ERPs through ear-EEG have been possible despite smaller amplitudes (as predicted within the simulations within this paper). A likely reason for this is the lower noise amplitudes within the ear-EEG, where noise originated both from untargeted brain signals and other endogenous sources such as eye movements and muscle activity. This theory is backed by the simulated examples of decreases in both signal and noise amplitude in ear-EEG in this paper. Indeed, an approximately proportional scaling of EEG and artifact sensitivities is observed for both unilateral and bilateral montages with each ocular artifact (compare the analysis of inter-electrode distance and sensitivity in Figure 6D with the ocular artifact channel sensitivities and topographical

plots in Figures 7A–F). Such results support the use of ear-EEG in the detection of neural activity from regions further away from the ear, despite lower signal amplitude.

Limitations and future work

The first limitation of the presented study is the absence of absolute sensitivity predictions (predictions of the amplitude of ear-EEG recordings). However, the amplitude of EEG signals is likely to change as a result of many factors, such as the skin-electrode contact quality or the size of the group of neurons recruited for the response. In the absence of absolute sensitivity predictions, the approach adopted within this study focused on relative differences between various channels. In this way, meaningful comparisons were drawn without requiring knowledge of absolute sensitivity values. Relative amplitudes are also useful in practice, since, regardless of the absolute amplitude of the response, the alignment between model predictions and measured relative amplitude between two channels can serve as an indicator for reliable EEG recordings. Nevertheless, absolute predictions could be made possible through calibration of the model with experimental measurements.

For a generalizable EEG model, geometry resembling that of the population average is desirable, since the aim of the modeling is to provide results that can be applied to most subjects. However, average head geometry data including tissues such as the eyes and ears, which are required for neural and ocular ear-EEG sensitivity modeling, does not exist. Therefore, the present volume conductor model was built using single subject head geometry data from the widely accepted MIDA model (Iacono et al., 2015). In order to mitigate the use of subject specific geometry, the analysis conducted in this paper reflected that of other generalizable EEG models, which focus on large-scale, general variations in geometry that are expected to be shared between large portions of the population. Moreover, considering that (i) the current model was based on the anatomical geometry of a subject with no known physical abnormalities, and (ii) inter-subject anatomical variability is sufficient to substantially limit the generalizability of average anatomical geometry, the presently used single-subject geometry is likely to suffice in producing generalizable results, provided that the interpretation of the results are within the bounds of generalizable modeling.

A 64-channel scalp EEG configuration was considered in this study. The scalp electrodes were placed according to the standard 10–20 system. We compared the source-sensor relationship of this scalp montage to that of the ear-EEG montage. This particular arrangement of scalp EEG electrodes was selected because of its high adoption rate in research, its relatively high density, and its wide spatial distribution across the scalp. However, it does not include many electrodes around the ears. Therefore, the reader should note that other scalp montages which contain a higher density of electrodes around the ear regions would likely produce sensitivity profiles more similar to that of ear-EEG for temporal lobe regions, and could probably perform with less, or even no signal loss (relative to ear-EEG) in temporal regions (Figures 5E–H).

A systematic analysis of the sensitivity to source orientation, position, and distance, such as that presented in Meiser et al. (2020) for cEEGrid, has not been the aim of the present study, which does not consider the analysis of the associated variations in source characteristics. One way in which such an analysis could be achieved with the current model is, for each of the 990 dipoles simulated in this study, to test varying orientations. Since few assumptions about the exact cortical folding structure of the brain would be made when employing this method, the absence of subject specific anatomy would be partially mitigated, enabling the model to maintain a level of generalizability.

In principle, the presented volume conductor modeling framework could be used to investigate the sensitivity of wearable EEG montages to muscle artifacts, for example by placing current sources in the locations from which EMG signals originate. An example of accurate volume conductor modeling of muscles has been provided by Pereira Botelho et al. (2019). The reader should be aware that the modeling method employed within Pereira Botelho et al. (2019) utilizes both the muscle fiber geometry and motor unit activation patterns for the muscle of interest, in addition to a modified current-source modeling approach which exploits the reciprocity theorem (Plonsey, 1963; Rush and Driscoll, 1969). Such methods were employed to gain accurate predictions, and in an efficient manner (the reciprocity theorem applied to single channel predictions enables reduced computation times for a large number of sources). Therefore, while EMG simulation within the current framework is feasible, more detailed current modeling than that which is shown here might be required for meaningful results.

For artifacts that do not arise as a result of an internal source of electric field, for example motion and external field artifacts, the presented modeling framework would require modification, such that the relevant physics is incorporated into the model. An advantage of simulating through COMSOL is that an existing model can be adapted and used in simulations of various physics domains, for example mechanics or electric fields and circuits. As such, it is possible that the presented model could be modified to incorporate simulations of head and electrode mechanics during motion, or external field interference.

Conclusion

Novel insights into the ear-EEG source sensor relationship for both neural and ocular sources have been provided, through comparisons of single channel and montage sensitivity profiles for ear and scalp EEG. The results have provided conclusive evidence for the use of ear-EEG in applications concerning the monitoring of neural activity originating from within the temporal lobes, for both unilateral and bilateral montages of ear EEG, while evidence has been provided for equal SNR

between ear-EEG and scalp EEG in the presence of ocular artifacts. The reported results could also be used as a reference for various neural and ocular sources, supporting both existing and prospective experimental ear-EEG studies. Future work will look to exploit the utility of the presented physics modeling to provide further insight into the sensitivity of ear-EEG to both neural and a variety of artifact sources.

Data availability statement

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

Ethics statement

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

Neural source modeling was devised and analyzed by MY and conducted by MY and MT with supervision from DM. Ocular source modeling was devised, conducted, and analyzed by MY and MT with supervision from DM. The manuscript was prepared by MY and MT and edited by all authors. All authors contributed to the article and approved the submitted version.

Funding

MY was supported by the Racing Foundation Grant Number: 285/2018, the USSOCOM Grant Number: EESB P85655, and the MURI/EPSRC Grant Number: EP/P008461. MT was supported by the UKRI CDT in AI for Healthcare (http://ai4health.io) (Grant Number: P/S023283/1). DM was partially supported by the USSOCOM Grant Number: EESB P85655. All funder's have supported both the research and publication for this paper.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.997377/full#supplementary-material

References

Abiri, R., Borhani, S., Sellers, E. W., Jiang, Y., and Zhao, X. (2019). A comprehensive review of EEG-based brain–computer interface paradigms. *J. Neural Eng.* 16, e011001. doi: 10.1088/1741-2552/aaf12e

Athavipach, C., Pan-Ngum, S., and Israsena, P. (2019). A wearable in-ear EEG device for emotion monitoring. *Sensors* 19, 4014. doi: 10.3390/s19184014

Bech Christensen, C., Hietkamp, R. K., Harte, J. M., Lunner, T., and Kidmose, P. (2018). Toward EEG-assisted hearing aids: Objective threshold estimation based on ear-EEG in subjects with sensorineural hearing loss. *Trends Hear.* 22, 2331216518816203. doi: 10.1177/2331216518816203

BESA (2022). *BESA: Brain Electrical Source Analysis*. Available online at: www.besa.de (accessed July 1, 2022).

Casson, A. J. (2019). Wearable EEG and beyond. *Biomed. Eng. Lett.* 9, 53–71. doi: 10.1007/s13534-018-00093-6

Christensen, C. B., Harte, J. M., Lunner, T., and Kidmose, P. (2017). Ear-eeg-based objective hearing threshold estimation evaluated on normal hearing subjects. *IEEE Trans. Biomed. Eng.* 65, 1026–1034. doi: 10.1109/TBME.2017.2737700

Coburn, K. L., and Moreno, M. A. (1988). Facts and artifacts in brain electrical activity mapping. *Brain Topogr.* 1, 37–45. doi: 10.1007/BF01129338

COMSOL (2022). *COMSOL Multiphysics®. v. 6.0*. Stockholm: COMSOL AB. Available online at: www.comsol.com (accessed July 1, 2022).

Gabriel, C. (1996). *Compilation of the Dielectric Properties of Body Tissues at RF and Microwave Frequencies. Technical Report*. London: King's College London, Department of Physics.

Goverdovsky, V., Von Rosenberg, W., Nakamura, T., Looney, D., Sharp, D. J., Papavassiliou, C., et al. (2017). Hearables: Multimodal physiological in-ear sensing. *Sci. Rep.* 7, 6948. doi: 10.1038/s41598-017-06925-2

Gratton, G. (1998). Dealing with artifacts: The EOG contamination of the event-related brain potential. *Behav. Res. Methods Instr. Comput.* 30, 44–53. doi: 10.3758/BF03209415

Grech, R., Cassar, T., Muscat, J., Camilleri, K. P., Fabri, S. G., Zervakis, M., et al. (2008). Review on solving the inverse problem in EEG source analysis. *J. Neuroeng. Rehabil.* 5, 1–33. doi: 10.1186/1743-0003-5-25

Hasgall, P. A., Di Gennaro, F., Baumgartner, C., Neufeld, E., Lloyd, B., Gosselin, M. C., et al. (2022). *IT'IS Database for Thermal and Electromagnetic Parameters of Biological Tissues: Version 4.1*. Available online at: itis.swiss/database

Iacono, M. I., Neufeld, E., Akinnagbe, E., Bower, K., Wolf, J., Vogiatzis Oikonomidis, I., et al. (2015). MIDA: A multimodal imaging-based detailed anatomical model of the human head and neck. *PLoS ONE* 10, e0124126. doi: 10.1371/journal.pone.0124126

Joyce, C. A., Gorodnitsky, I. F., and Kutas, M. (2004). Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. *Psychophysiology* 41, 313–325. doi: 10.1111/j.1469-8986.2003.00141.x

Kappel, S. L., Christensen, C. B., Mikkelsen, K. B., and Kidmose, P. (2016). "Reference configurations for ear-EEG steady-state responses," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Orlando, FL, 5689–5692.

Kappel, S. L., and Kidmose, P. (2017). "High-density ear-EEG," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Jeju, 2394–2397.

Kappel, S. L., Looney, D., Mandic, D. P., and Kidmose, P. (2017). Physiological artifacts in scalp EEG and ear-EEG. *Biomed. Eng. Onl.* 16, 103. doi: 10.1186/s12938-017-0391-2

Kappel, S. L., Makeig, S., and Kidmose, P. (2019). Ear-EEG forward models: Improved head-models for ear-EEG. *Front. Neurosci.* 13, 943. doi: 10.3389/fnins.2019.00943

Kidmose, P., Looney, D., Ungstrup, M., Rank, M. L., and Mandic, D. P. (2013). A study of evoked potentials from ear-EEG. *IEEE Trans. Biomed. Eng.* 60, 2824–2830. doi: 10.1109/TBME.2013.2264956

Lins, O. G., Picton, T. W., Berg, P., and Scherg, M. (1993a). Ocular artifacts in EEG and event-related potentials I: Scalp topography. *Brain Topogr.* 6, 51–63. doi: 10.1007/BF01234127

Lins, O. G., Picton, T. W., Berg, P., and Scherg, M. (1993b). Ocular artifacts in recording EEGs and event-related potentials II: Source dipoles and source components. *Brain Topogr.* 6, 65–78. doi: 10.1007/BF01234128

Looney, D., Kidmose, P., Park, C., Ungstrup, M., Rank, M. L., Rosenkranz, K., et al. (2012). The in-the-ear recording concept: User-centered and wearable brain monitoring. *IEEE Pulse* 3, 32–42. doi: 10.1109/MPUL.2012.2216717

Looney, D., Park, C., Kidmose, P., Rank, M. L., Ungstrup, M., Rosenkranz, K., et al. (2011). An in-the-ear platform for recording electroencephalogram. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2011, 6882–6885. doi: 10.1109/IEMBS.2011.6091733

Ma, J., Tao, P., Bayram, S., and Svetnik, V. (2012). Muscle artifacts in multichannel EEG: Characteristics and reduction. *Clin. Neurophysiol.* 123, 1676–1686. doi: 10.1016/j.clinph.2011.11.083

Martinek, J., Stickler, Y., Reichel, M., Mayr, W., and Rattay, F. (2008). A novel approach to simulate Hodgkin–Huxley-like excitation with COMSOL Multiphysics. *Artif. Organs* 32, 614–619. doi: 10.1111/j.1525-1594.2008.00611.x

Meiser, A., Tadel, F., Debener, S., and Bleichner, M. G. (2020). The sensitivity of ear-EEG: Evaluating the source-sensor relationship using forward modeling. *Brain Topogr.* 33, 665–676. doi: 10.1007/s10548-020-00793-2

Mikkelsen, K. B., Kappel, S. L., Mandic, D. P., and Kidmose, P. (2015). EEG recorded from the ear: Characterizing the ear-EEG method. *Front. Neurosci.* 9, 438. doi: 10.3389/fnins.2015.00438

Mikkelsen, K. B., Villadsen, D. B., Otto, M., and Kidmose, P. (2017). Automatic sleep staging using ear-EEG. *Biomed. Eng. Onl.* 16, 1–15. doi: 10.1109/EMBC.2016.7591789

Mosher, J. C., Lewis, P. S., and Leahy, R. M. (1992). Multiple dipole modeling and localization from spatio-temporal MEG data. *IEEE Trans. Biomed. Eng.* 39, 541–557. doi: 10.1109/10.141192

Muthukumaraswamy, S. (2013). High-frequency brain activity and muscle artifacts in MEG/EEG: A review and recommendations. *Front. Hum. Neurosci.* 7, 138. doi: 10.3389/fnhum.2013.00138

Nakamura, T., Alqurashi, Y. D., Morrell, M. J., and Mandic, D. P. (2018). "Automatic detection of drowsiness using in-ear EEG," In *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN)*. Rio de Janeiro, 1–6.

Nakamura, T., Goverdovsky, V., and Mandic, D. P. (2017a). In-ear EEG biometrics for feasible and readily collectable real-world person authentication. *IEEE Trans. Inform. For. Secur.* 13, 648–661. doi: 10.1109/TIFS.2017.2763124

Nakamura, T., Goverdovsky, V., Morrell, M. J., and Mandic, D. P. (2017b). Automatic sleep monitoring using ear-EEG. *IEEE J. Transl. Eng. Health Med.* 5, 1–8. doi: 10.1109/JTEHM.2017.2702558

Noachtar, S., and Rémi, J. (2009). The role of EEG in epilepsy: A critical review. *Epilepsy Behav.* 15, 22–33. doi: 10.1016/j.yebeh.2009.02.035

Oliveira, A. S., Schlink, B. R., Hairston, W. D., König, P., and Ferris, D. P. (2016). Induction and separation of motion artifacts in EEG data using a mobile phantom head device. *J. Neural Eng.* 13, e036014. doi: 10.1088/1741-2560/13/3/036014

Pelot, N. A., Thio, B. J., and Grill, W. M. (2018). Modeling current sources for neural stimulation in COMSOL. *Front. Comput. Neurosci.* 12, 40. doi: 10.3389/fncom.2018.00040

Pereira Botelho, D., Curran, K., and Lowery, M. M. (2019). Anatomically accurate model of EMG during index finger flexion and abduction derived from diffusion tensor imaging. *PLoS Comput. Biol.* 15, e1007267. doi: 10.1371/journal.pcbi.1007267

Plonsey, R. (1963). Reciprocity applied to volume conductors and the ECG. *IEEE Trans. Biomed. Electr.* 10, 9–12. doi: 10.1109/TBMEL.1963.4322775

Richer, N., Downey, R. J., Nordin, A. D., Hairston, W. D., and Ferris, D. P. (2019). "Adding neck muscle activity to a head phantom device to validate mobile EEG muscle and motion artifact removal," In *Proceedings of the 9th International IEEE/EMBS Conference on Neural Engineering (NER)*. San Francisco, CA, 275–278.

Roy, R. N., Charbonnier, S., and Bonnet, S. (2014). Eye blink characterization from frontal EEG electrodes using source separation and pattern recognition algorithms. *Biomed. Sign. Process. Contr.* 14, 256–264. doi: 10.1016/j.bspc.2014.08.007

Rush, S., and Driscoll, D. A. (1968). Current distribution in the brain from surface electrodes. *Anesth. Analg.* 47, 717–723.

Rush, S., and Driscoll, D. A. (1969). EEG electrode sensitivity- an application of reciprocity. *IEEE Trans. Biomed. Eng.* 16, 15–22. doi: 10.1109/TBME.1969.4502598

Sarvas, J. (1987). Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Phys. Med. Biol.* 32, 11. doi: 10.1088/0031-9155/32/1/004

Schulman-Galambos, C., and Galambos, R. (1979). Brain stem evoked response audiometry in newborn hearing screening. *Archiv. Otolaryngol.* 105, 86–90. doi: 10.1001/archotol.1979.00790140032006

Seibt, O., Truong, D., Khadka, N., Huang, Y., and Bikson, M. (2019). Computational finite element method (FEM) forward modeling workflow for transcranial direct current stimulation (tDCS) current flow on MRI-derived head: Simpleware and comsol multiphysics tutorial. *bioRxiv* 2019, 704940. doi: 10.1101/704940

Symeonidou, E.-R., Nordin, A. D., Hairston, W. D., and Ferris, D. P. (2018). Effects of cable sway, electrode surface area, and electrode mass on electroencephalography signal quality during motion. *Sensors* 18, 1073. doi: 10.3390/s18041073

Von Rosenberg, W., Chanwimalueang, T., Goverdovsky, V., Looney, D., Sharp, D., and Mandic, D. P. (2016). Smart helmet: Wearable multichannel ECG and EEG. *IEEE J. Transl. Eng. Health Med.* 4, 2609927. doi: 10.1109/JTEHM.2016.2609927

Webster, J. G. (2009). *Medical instrumentation: Application and Design*. New York, NY: John Wiley & Sons.

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

Yarici, M. C., Davies, H. J., Nakamura, T., Williams, I., and Mandic, D. P. (2021). "Hearables: In-ear multimodal brain computer interfacing," in *Brain-Computer Interface Research*, ed. B. Allison (Berlin/Heidelberg: Springer), 79–87.

Zibrandtsen, I., Kidmose, P., Christensen, C., and Kjaer, T. (2017). Ear-EEG detects ictal and interictal abnormalities in focal and generalized epilepsy–A comparison with scalp EEG monitoring. *Clin. Neurophysiol.* 128, 2454–2461. doi: 10.1016/j.clinph.2017.09.115

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 01 February 2023
DOI 10.3389/fnins.2023.987578

OPEN ACCESS

EDITED BY
Yael Hanein,
Tel Aviv University, Israel

REVIEWED BY
Han-Jeong Hwang,
Korea University, Republic of Korea
Emidio Pistilli,
West Virginia University, United States

*CORRESPONDENCE
Preben Kidmose
✉ pki@ece.au.dk

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

RECEIVED 06 July 2022
ACCEPTED 09 January 2023
PUBLISHED 01 February 2023

CITATION
Tabar YR, Mikkelsen KB, Shenton N, Kappel SL, Bertelsen AR, Nikbakht R, Toft HO, Henriksen CH, Hemmsen MC, Rank ML, Otto M and Kidmose P (2023) At-home sleep monitoring using generic ear-EEG. *Front. Neurosci.* 17:987578. doi: 10.3389/fnins.2023.987578

At-home sleep monitoring using generic ear-EEG

Yousef R. Tabar[1], Kaare B. Mikkelsen[1], Nelly Shenton[2], Simon L. Kappel[1], Astrid R. Bertelsen[2], Reza Nikbakht[2], Hans O. Toft[2], Chris H. Henriksen[2], Martin C. Hemmsen[2], Mike L. Rank[2], Marit Otto[3] and Preben Kidmose[1]*

[1]Department of Electrical and Computer Engineering, Aarhus University, Aarhus, Denmark, [2]T&W Engineering A/S, Allerød, Denmark, [3]Department of Clinical Neurophysiology, Aarhus University Hospital, Aarhus, Denmark

Introduction: A device comprising two generic earpieces with embedded dry electrodes for ear-centered electroencephalography (ear-EEG) was developed. The objective was to provide ear-EEG based sleep monitoring to a wide range of the population without tailoring the device to the individual.

Methods: To validate the device ten healthy subjects were recruited for a 12-night sleep study. The study was divided into two parts; part A comprised two nights with both ear-EEG and polysomnography (PSG), and part B comprised 10 nights using only ear-EEG. In addition to the electrophysiological measurements, subjects filled out a questionnaire after each night of sleep.

Results: The subjects reported that the ear-EEG system was easy to use, and that the comfort was better in part B. The performance of the system was validated by comparing automatic sleep scoring based on ear-EEG with PSG-based sleep scoring performed by a professional trained sleep scorer. Cohen's kappa was used to assess the agreement between the manual and automatic sleep scorings, and the study showed an average kappa value of 0.71. The majority of the 20 recordings from part A yielded a kappa value above 0.7. The study was compared to a companioned study conducted with individualized earpieces. To compare the sleep across the two studies and two parts, 7 different sleeps metrics were calculated based on the automatic sleep scorings. The ear-EEG nights were validated through linear mixed model analysis in which the effects of equipment (individualized vs. generic earpieces), part (PSG and ear-EEG vs. only ear-EEG) and subject were investigated. We found that the subject effect was significant for all computed sleep metrics. Furthermore, the equipment did not show any statistical significant effect on any of the sleep metrics.

Discussion: These results corroborate that generic ear-EEG is a promising alternative to the gold standard PSG for sleep stage monitoring. This will allow sleep stage monitoring to be performed in a less obtrusive way and over longer periods of time, thereby enabling diagnosis and treatment of diseases with associated sleep disorders.

KEYWORDS

electroencephalography, sleep monitoring, ear-EEG, long-term sleep monitoring, home recording

1. Introduction

Lack of sleep and poor sleep quality is a grand societal challenge. Poor sleep quality has a negative impact on health, the feeling of wellbeing, quality of life and on human cognitive performance. This has large negative consequences for society, productivity, and economy. Recently, sleep researchers and physicians of sleep medicine have emphasized the importance of sleep on human's health (Naismith et al., 2011; Galbiati et al., 2019; Stefani and Högl, 2020). However, the development within the field of sleep monitoring has not evolved much beyond today's gold standard, polysomnography (PSG). As its name describes, PSG is a method that encompasses multiple modalities to describe the sleep. These measurements are typically recorded in sleep clinics and include recordings of brain activity [electroencephalography (EEG)], eye movements [electrooculography (EOG)], muscle [electromyography, (EMG)] and heart activity [electrocardiography (ECG)] (Berry et al., 2012). The PSG's range of sensors and wiring means that PSG monitoring is discomfortable and in consequence has a negative impact on the sleep. Furthermore, sleep monitoring often takes place in a sleep clinic instead of in the patient's home environment and sleeping in an unfamiliar environment also has a negative impact on sleep. Although patients may to some extent become accustomed to the equipment and unfamiliar environment, so that the effects become less of a problem over time, the phenomenon is well-known and are called the "first night effect" and implies e.g., a reduced total sleep time, a decrease in sleep efficiency, and delayed and decreased REM sleep (Agnew et al., 1966). These aspects, together with the dependence on health professionals and the significant cost associated has limited the PSG's use in long-term sleep monitoring. In this light, there is a need for an efficient, comfortable, and easy to use device for monitoring sleep at home.

Alternative sleep assessment methods such as sleep diaries and actigraphy are currently used in many sleep studies and clinical investigations (Sadeh, 2015; Zhu et al., 2018). Unfortunately, the amount of information gained from these methods is limited compared to that of PSG. They are therefore mainly used in parallel with PSG (Gaina et al., 2004). In recent years, several simple monitoring systems have been introduced for sleep assessment. These devices rely on fewer sensors (mostly dry EEG electrodes) to increase comfort and ease of use. The Dreem headband (Arnal et al., 2019), the forehead mounted Prodigy device (Younes et al., 2017) and ear-centered electroencephalography (ear-EEG) device (Mikkelsen et al., 2019; Nakamura et al., 2019) are some examples. While comfort and ease of use make these devices ideal for sleep assessment, they are still to be validated in vaster studies before they can be implemented in the clinic.

Ear-centered electroencephalography was first introduced in 2011 (Looney et al., 2011). The original aim was to provide a more comfortable and affordable solution for several neurophysiological problems at a negligible performance cost. Ear-EEG is a method in which EEG signals are recorded from electrodes in or around the ear. A large variety of different solutions have been proposed, including electrodes placed solely around-the-ear (Bleichner and Debener, 2017), electrodes on customized earpieces (Mikkelsen et al., 2015) and more generic type earpieces (Goverdovsky et al., 2015). One of the most advanced methods are based on dry-contact electrodes embedded on individualized earpieces made of soft silicone (Kappel et al., 2018). The first ear-EEG sleep assessment study was performed in 2019 (Mikkelsen et al., 2019, 2021b; Tabar et al., 2020, 2021) using custom made earplugs and a commercial amplifier. In the current study we deployed a recent advancement in the development of a comfortable, generic, and ready to use setup for sleep assessment, without the sacrifice of performance. In this article, we present an at-home sleep monitoring setup with generic earpieces and a proprietary amplifier for ear-EEG sleep assessment in healthy people.

The focus of this article is on the comfort of the presented ear-EEG device, the data quality of the recordings and the resulting hypnograms. First, we present the generic earpiece design. Then, we introduce our custom EEG amplifier. Next, we present the feedback on the comfort of the earpieces. Finally, the data quality and hypnograms are presented.

2. Materials and methods

2.1. Experimental setup

2.1.1. The generic earpieces

A prerequisite for this study was the development of a generic earpiece accommodating ear-EEG based sleep monitoring to a wide range of the population. The generic earpiece should provide a reliable and robust contact between the body and the electrodes embedded in the earpiece, and at the same time the earpiece should fit most human ears, be easy to use and be comfortable to sleep with. To begin the design process, we studied the anatomy of the human ear, to identify sizes, curves, and anatomical landmarks. This was achieved by examination of a large number of 3D scans of human ears and through review of existing literature in the domain (Toivonen et al., 2002; Lee et al., 2018; Modabber et al., 2018). This was used as the first input to the design process. The earpiece design process was iterative and in each step in the process we evaluated both the comfort, the ease of use and the EEG signal quality.

There is a very large variety in anatomical shape and sizes of human ears, and to accommodate most ears, it was necessary to design four different earpieces. All four earpieces had the same basic shapes but varied in sizes. The earpieces consisted of three main features: an ear canal part, a tail, and a main body. The ear canal part was given a tulip-like shape to ensure easy insertion in the ear canal, while still filling out the cross section of the ear canal. The ear canal was designed in three different sizes, equivalent to cross sectional diameters of 7, 8, and 9 mm. Essential to the comfort is how deep the earpiece goes into the ear canal and how well it fits the anatomy of the ear. In our design the earpiece was relative shallow and was not going more than 5–6 mm into the ear canal. The tail was designed to keep the earpiece in place by applying pressure on the outer edge of concha. The main body acted as a connector between the two other parts and formed a smooth transition between the concha and the outer part of the ear canal. Additionally, the main body served as a strain relief for the cable to minimize motion artifact from cable pulling. Both the main body and tail were available in two sizes. Eventually we selected four size combinations of the three main features to be used in the study.

The comfort of an earpiece is related to a wide range of factors including mechanical properties of the earpiece material, the number and placement of electrodes, and the ergonomic design. Our experience was that the earpieces should be made of a soft and

compliant material. Thus, the earpieces were molded in a soft silicone material (Detax Software 2.0, DETAX GmbH, Germany).

Regarding the number and position of the electrodes, for sleep monitoring the most important factor is to have a good cross-ear derivative; this aspect has been investigated in detail in Mikkelsen et al. (2021a). Thus, the number of electrodes in each ear is a trade-off between comfort and redundancy; increasing the number of electrodes gives a higher degree of redundancy at the cost of lower comfort. In our design we decided to use two recording electrodes in each ear, which is significantly lower than the six electrodes in each ear used in previous studies (Mikkelsen et al., 2017, 2019). The electrodes were placed where the earpiece seemed to apply the largest pressure toward the body, because this is believed to give the best contact between the electrode and the body.

Finally, the shielded cables where assembled in a form-stable and arc-formed tube guiding the cables around the superior part of the Helix of the ear, see **Figure 1**. This further reduces the effect of cable movement and make the earpiece more discrete. Electrodes were embedded in the earpieces; the remaining electronics were placed in a box outside the ear to ensure good comfort.

2.1.2. Custom-designed EEG-amplifier and electrodes

The EEG amplifier was a 4-channel "EEG-to-digital converter" application specific integrated circuit (ASIC) specifically developed for ear-EEG measurements. The ASIC was optimized for high input impedance, high common mode rejection, low current noise, and low power consumption. The ASIC was a revised version of the design described in Zhou et al. (2016); it was extended from 2 to 4 channels and a digital control block was included to make it easier to store the data. The EEG amplifier was connected to the electrodes embedded in the earpieces. Two electrodes in each earpiece were used as recording electrodes, one electrode in the left earpiece was used as reference electrode and one electrode in each earpiece was used as ground. The sleep study measurements were all recorded at 250 Hz sampling rate and with 14 bit resolution.

The electrodes were made of Titanium with a porous coating of Iridium Oxide at the contact surface, see (Kappel et al., 2018) for a detailed description and characterization. The electrodes were circular with a diameter of 2.6 mm and with a slight concave shape, whereby the electrodes protruded slightly from the surface of the earpieces. Each electrode was connected to the amplifier with a Ø 0.53 mm coaxial cable. The cable shielding was extended all the way to the back side of the electrode. The shield was actively driven by a unity gain amplifier in the ASIC.

2.2. Sleep recordings

This study was approved by the Central Denmark Region Committees on Biomedical Research Ethics (Ref. nr. 1-10-72-13-20) as well as the Danish Medicines Agency (ref. nr. 2020012619). Written informed consent was obtained from the participants prior to participation. 10 subjects (4 f, 6 m) participated in this study. The ages of the subjects ranged between 22 and 35, with a mean of 27.4 years. Participants were screened for hearing loss, sleep disorders, neurological disorders, bruxism, pregnancy, drug usage, allergies, chronic pain, and sleep apnea. Each participant attended an earpiece fitting session prior to the recordings. During this session, the earpieces with the best fit to the participant's ears was determined by visual inspection of the EEG signal, and the participants were trained in mounting the earpieces themselves. The subjects were instructed to put on the earpieces and start recording whenever they wanted to go to bed and stop the recording when they wake up. They were free to spend any time in bed before sleeping. The participants were asked to fill a sleep diary during the recording and a questionnaire immediately after the wake up.

The study was divided in two parts, Part A and Part B. In Part A, participants slept two nights with the partial PSG (EEG, EOG, and EMG electrodes) and ear-EEG setup. Please refer to Mikkelsen et al. (2019) for more details about the partial PSG. Following two successful recordings in Part A, the study proceeded to Part B in which the participants recorded ten full nights using only the ear-EEG setup. The recordings were performed in the participants' own home. In Part A, participants visited the laboratory on the day of the recording to get the partial PSG mounted. The participants mounted the earpieces themselves just before the start of the recording. Ear-EEG and partial PSG were recorded using different data acquisition devices. The participants were asked to press a trigger button on both devices simultaneously for signal alignment. The partial PSG recordings were manually scored by a trained professional sleep technologist according to the AASM manual for the Scoring of Sleep and Associated Events (Berry et al., 2012).

In the following sections the recordings described above will be referred to as sG, a dense form referring to the study with generic earpieces. Specifically, sG will refer to the complete dataset, and sG.A and sG.B will refer to the dataset related to Part A and Part B, respectively. In addition to this dataset, we also used a companioned dataset from a previous ear-EEG sleep monitoring study (Mikkelsen et al., 2019, 2022) in which custom-made earpieces and a commercial amplifier were used. This dataset will be referred to as sC, a dense form referring to the study with custom earpieces. sC was structured in the same way as sG, but with 20 subjects recorded four nights with both partial PSG and ear-EEG in part A (referred to as sC.A) and 10 subjects recorded 12 nights with only ear-EEG in part B (referred to as sC.B). A summary of the datasets used in this article is presented in **Figure 2**. It is important to note that in the current study, sC was only used for training the sleep scoring classifier and for statistical comparisons.

2.3. Assessment of sleep quality, comfort and ease-of-use

The participants reported perceived sleep quality, comfort and ease-of-use of the device by rating six questions on a Likert scale. The six questions and the corresponding Likert scale can be seen in **Figure 3**. The participants were instructed to fill out the questionnaire just after they woke up in the morning. Inter, intra subject variation values were computed as the standard deviation of the ratings between the subjects and within the subjects.

2.4. Signal pre-processing and channel selection

A signal pre-processing pipeline was applied to the recorded ear-EEG data. Each channel was bandpass filtered (0.1–100 Hz),

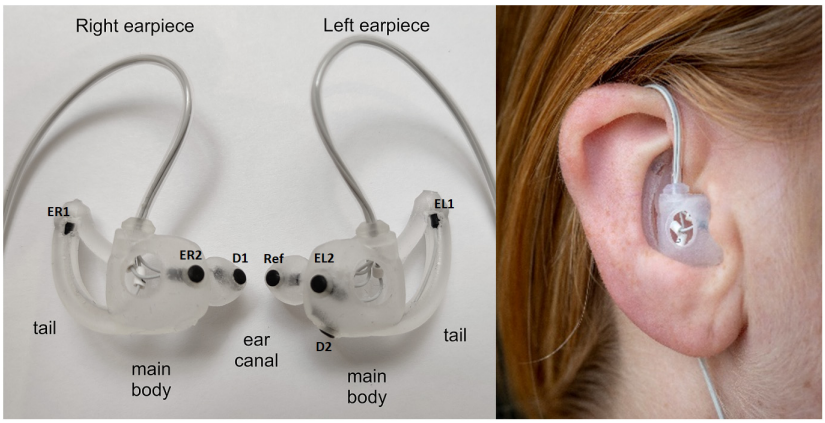


FIGURE 1
(Left) Generic earpieces with mounted electrodes. The placement of the three main features (ear canal, tail, and main body) are specified for each earpiece. The generic earpiece was molded in soft silicone in one piece to ensure good comfort. Cable relief was incorporated in the design to reduce movement artifacts from cable pulling. Also, a formed tube was designed to guide the cables behind the ear to further reduce cable movement, to increase comfort and to make the earpiece more discrete. EL1, EL2, ER1, and Er2: data electrodes, D1 and D2 ground electrodes, ref: reference electrode. **(Right)** Earpiece mounted in the ear.

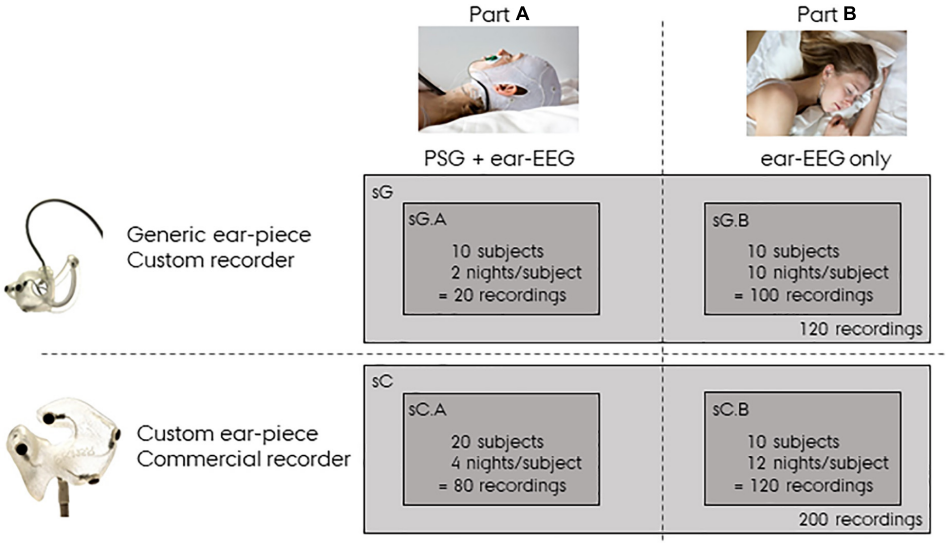


FIGURE 2
Overview of the data. sG refers to data from the study with *generic* earpieces, and sC to data from the study with *custom* earpieces. Both datasets have a Part A and a Part B. Part A were recorded with both polysomnography (PSG) and ear-centered electroencephalography (ear-EEG), whereas Part B were recorded from ear-EEG only.

and notch filtered at 50 and 100 Hz to suppress power line noise. Artifacts were identified and removed following several steps. For the sG dataset we observed periodic noise spikes with a period of 200 ms. The noise was related to internal communication in custom developed amplifier and the severity of the noise increased with electrode impedance, i.e., the higher the electrode impedance the more prone the channel was to the device induced noise. The noise was detected by an algorithm looking for spikes with a repetition rate of 200 ms. If a spike was detected it was removed and the signal was interpolated based on the clean signal in the neighboring samples.

Spikes of short duration and high amplitude were also detected and removed. These spikes are usually due to small changes in the electrode-skin connection. Poor electrode-skin connection can also lead to dominant high frequency noise in the signal. Long periods (>30 s) with such a noise were detected by thresholding the high frequency power. Finally, any sample with an absolute value greater than 350 μV were rejected. Exclusion of unsuccessful recordings from the dataset was performed if more than 30% of the signal was noisy or if the duration of the recording was less than 5 h.

In the **Supplementary material**, we have supplied figures showing the effect of both spike removal and general noise removal.

Finally, the four data channels were combined to construct a single channel signal. To extract this signal, a channel selection method based on root mean square (RMS) was employed. The method relied on the idea that noisy signals tend to yield higher RMS values. Accordingly, for every 30 s epoch, RMS values were computed for any possible cross-ear combination of the channels. The combinations were constructed by using channels from right ear

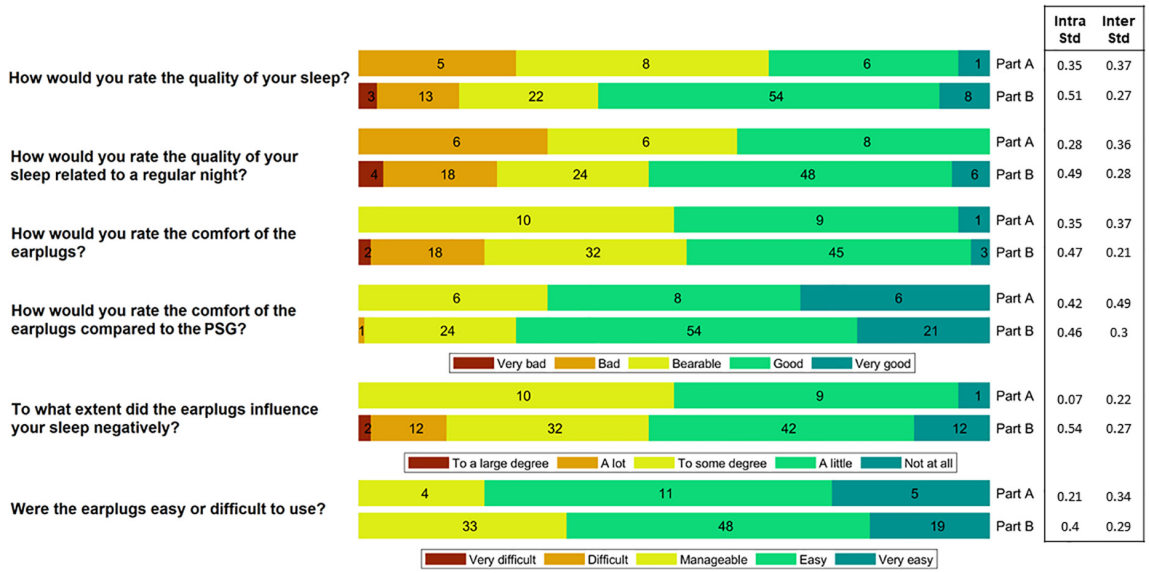


FIGURE 3
The participants rated the comfort and ease-of-use of the device after each night. In Part A, they slept wearing both the polysomnography (PSG) setup and the earpieces, whereas in Part B they only wore the earpieces. This figure shows a summary of their ratings. The numbers in the bars reflect the number of responses (20 in total for Part A and 100 in total for Part B). Inter subject variation (Inter Std) and Intra subject variation (Intra Std) are shown for each question-part combination.

[ER1, ER2, and Avg(ER1,ER2)] referenced to the channels from left ear [EL1, EL2, Avg(EL1,EL2), and ref]. The channel combination that yielded that lowest RMS value was selected.

2.5. Automatic sleep scoring

Following the pre-processing step and the construction of a single channel signal, an automatic sleep scoring algorithm was applied to the resulting signal. The goal of the automatic sleep scoring was to assign a correct sleep stage (N1, N2, N3, REM, and wake) to every 30 s epoch.

First, the signal was segmented into 30 s long epochs. Then, a feature extraction step was applied to every epoch resulting in 84 features for each epoch. The feature set was selected to include time domain, frequency domain, Continuous Wavelet Transform (CWT) based, EMG proxy, EOG proxy, sleep event proxies, and non-linear features [adapted from Mikkelsen et al. (2019)]. The list of features is presented in **Supplementary Table 1**. Since sC and sG were recorded with different devices, the computed features were normalized using the normalize function in MATLAB. The normalization process was applied to each recording separately.

The epochs were then classified using a five-class random forest classifier consisting of 100 decision trees. This method has been used in several ear-EEG sleep scoring studies (Khademi et al., 2018; Mikkelsen et al., 2019; Tabar et al., 2021). The classification performance was measured using Cohen's kappa value (Cohen, 1960) which reflects the agreement between automatic and manual scoring. A "leave-one-subject-out" (LOSO) cross-validation was used to validate the results. This means that for each subject, the classifier was trained using only recordings from the remaining subjects. Since manual scoring was only available for the Part A recordings, the classification results were computed only for Part A.

2.6. Analysis of the effects of equipment, part, and subject

The Part B recordings were performed without PSG and therefore the performance of the automatic sleep scoring cannot be validated using Cohen's kappa value. Instead, we computed the following sleep metrics for all recordings:

$$REMfr = \frac{number\ of\ REM\ epochs}{total\ number\ of\ epochs} \quad (1)$$

$$N3fr = \frac{number\ of\ N3\ epochs}{total\ number\ of\ epochs} \quad (2)$$

$$SE\left(sleep\ efficiency\right) = \frac{sleep\ duration}{duration\ of\ the\ recording\ after\ first\ sleep\ epoch} \quad (3)$$

$$NREMtoNREM = \frac{number\ of\ NREM\ to\ NREM\ transitions}{total\ number\ of\ NREM\ epochs} \quad (4)$$

$$NREMtoREM = \frac{number\ of\ NREM\ to\ REM\ transitions}{total\ number\ of\ NREM\ epochs} \quad (5)$$

$$REMtoREM = \frac{number\ of\ REM\ to\ REM\ transitions}{total\ number\ of\ REM\ epochs} \quad (6)$$

$$REMtoNREM = \frac{number\ of\ REM\ to\ NREM\ transitions}{total\ number\ of\ REM\ epochs} \quad (7)$$

We investigated the effect of equipment (individualized vs. generic earpieces), part (PSG and ear-EEG vs. only ear-EEG) and

subject on each of these metrics using Linear Mixed Models (LMM). The objective was to investigate if differences in sleep characteristics could be explained by the equipment, part or inter-subject variation. For each sleep metric, a LMM was designed as [in Wilkinson notation (Wilkinson and Rogers, 1973)]:

$$Sleep\ metric \sim 1 + subject + part + study + (1 \mid recording) \quad (8)$$

Where equipment = 1, 2 (corresponding to individualized and generic earpieces), part = A, B (corresponding to combined PSG and ear-EEG, and only ear-EEG), and subject = 1, . . ., 20 were included as fixed effects and recording = 1, . . ., 16 / recording = 1, . . ., 12 was included in the model as a random effect. By fitting the models to the data, we examined the effect of each variable on the sleep metrics.

3. Results

Ten healthy participants (6 m/4 f) aged 27.4 ± 4.9 years were included in the study. From these 10 subjects, we collected 20 nights of combined PSG and ear-EEG and 100 nights of ear-EEG sleep recordings. Three recordings from Part A and nine recordings from Part B were rejected and repeated due to ear-EEG recording problems. The reasons for rejection of these recordings were: bad earpiece mounting (five recordings), low battery on recording device (four recordings) and earpiece fallen out of the ear during sleep (three recordings). Partial PSG recordings were not checked for noise. However, two recordings were repeated due to failure of the acquisition device.

3.1. Comfort and ease-of-use

Following each night, the participants rated the comfort and ease-of-use of the device by answering a questionnaire. A summary of the answers is illustrated in Figure 3. During the part of the study where sleep was assessed using both PSG and ear-EEG (Part A), 65% of the responses reported that sleep quality was "bad" or "bearable." In part B, using only the generic earpiece, 62% reported a "good" or "very good" sleep quality. This shift in perceived sleep quality fits well with their ratings of the comfort of the earplugs compared to the PSG setup, where 75% of the responses showed that the comfort was "good" or "very good." Inter and intra subject variation values are also presented in Figure 3. These values were observed to be generally low for all question-part combinations.

3.2. Data quality

The artifact rejection procedure described in section "2.4. Signal pre-processing and channel selection" led to rejection of on average 10.2% of the data. The proportion of each artifact rejection criterion was: device related noise: 3.2%, spikes 2.6%, high frequency: 3.9%, high amplitude: 0.5%. A summary of the pre-processing data rejection is presented in Figure 4.

The channel selection procedure described in section "2.4. Signal pre-processing and channel selection" was used to find the cross-ear combination with the least RMS among the accepted channels. Using this procedure, the average number of rejected epochs was 4.4%. The inter subject variation for data rejection was 0.027 for part A and 0.036 for part B.

3.3. Sleep scoring algorithm

The performance of the sleep scoring algorithm was validated using the sC.A and sG.A datasets. Different training and testing strategies were used to assess the sleep scoring, in which LOSO cross validation was used where applicable. sC.A included 80 recordings from 20 participants and sG.A included 20 recordings from 10 participants. The classifier was trained separately using sC.A, sG.A and a combination of sC.A and sG.A, and tested on both sC.A and sG.A. For simplicity, we called each of these cross-validation schemes XY, where X is the train set and Y is the test set, e.g., sCGsG means trained with sC.A and sG.A combined and tested on sG.A. A summary of the results of the different combinations is shown in Table 1. The confusion matrix for 5 class classification using the sCGsG method is presented in Figure 5.

While the average kappa value for sCsG was 0.68, it increased to 0.69 for the sGsG method. The number of recordings in the training set was only 18 in sGsG compared to 80 in sCsG. However, it still resulted in slightly better classification performance. The average kappa value further increased to 0.71 when both datasets were included in the training set in sCGsG. This is the highest kappa value we achieved for sG.A.

The right panel in Figure 6 shows the kappa values for each recording using the sCsG, sGsG and sCsG cross validation scheme. The kappa values of subjects 3 and 7 were conspicuously lower than for the other subjects. The scoring of the remaining 7 subjects resulted in a mean kappa well over 0.7 regardless of the training set. The three left panels in Figure 6 shows the distribution of kappa values for each cross-validation scheme. For the sGsG and sCGsG cross validation schemes the majority of the recordings have kappa values above 0.7. The average kappa value is shown for each method with a colored dashed line.

The proportion of decision trees voting for a given sleep stage can be interpreted as an estimate of the likelihood for that sleep stage. The output of the classifier is the sleep stage with the largest proportion of votes, and the proportion itself can be interpreted as a confidence measure of the classifier's decision. In a previous study (Mikkelsen et al., 2020), we observed that the mean value of the confidence measure across all epochs in a recording is a reliable estimate of the overall scoring performance. Thereby the confidence measure can be used to assess the sleep scoring performance for unlabeled data. To illustrate this relation between kappa and confidence the left panel in Figure 7 shows the confidence versus the kappa for the sG.A recordings. It is clear that the confidence values correlate positively with kappa values, which corroborate that the confidence is a reliable estimate of the kappa value. For these G.A recordings, the 25th, 50th, and 75th percentiles of the kappa values were 0.66, 0.72, and 0.76, respectively. In other words 75% of the recordings had a kappa value larger than 0.66, and the 25% best recordings had a kappa value larger than 0.76. The corresponding confidence values were 0.68, 0.70, and 0.71, respectively. The distribution of the confidence values for the sG.B recordings are shown in the second and third subfigure of Figure 7. In 58% of the Part B recordings, a confidence value above 25th percentile of Part A was achieved. This value was 42% for 50th percentile and 26% for 75th percentile. All subjects had at least 2 recordings with a confidence above the 25th percentile.

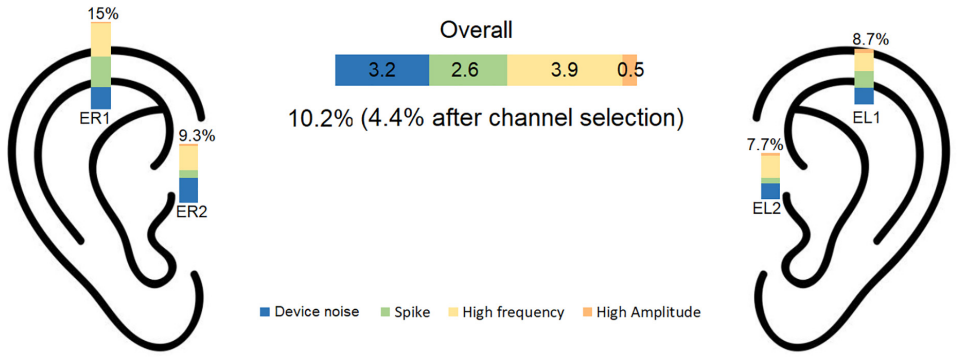


FIGURE 4
Percentage of rejected data is illustrated for each noise category for each electrode. The overall rejection percentage was 10.2% before and 4.4% after channel selection. EL1, left electrode 1; EL2, left electrode 2; ER1, right electrode 1; ER2, right electrode 2.

TABLE 1 The kappa values for each train-test pair together with the applied cross validation method.

Method	sCsC	sCsG	sGsG	sGsC	sCGsC	sCGsG
Training set	sC.A	sC.A	sG.A	sG.A	sC.A,sG.A	sC.A,sG.A
Test set	sC.A	sG.A	sG.A	sC.A	sC.A	sG.A
Cross validation	LOSO	–	LOSO	–	LOSO	LOSO
kappa	0.73	0.68	0.69	0.66	0.73	**0.71**

Highest performance on the sG.A dataset is marked in bold.

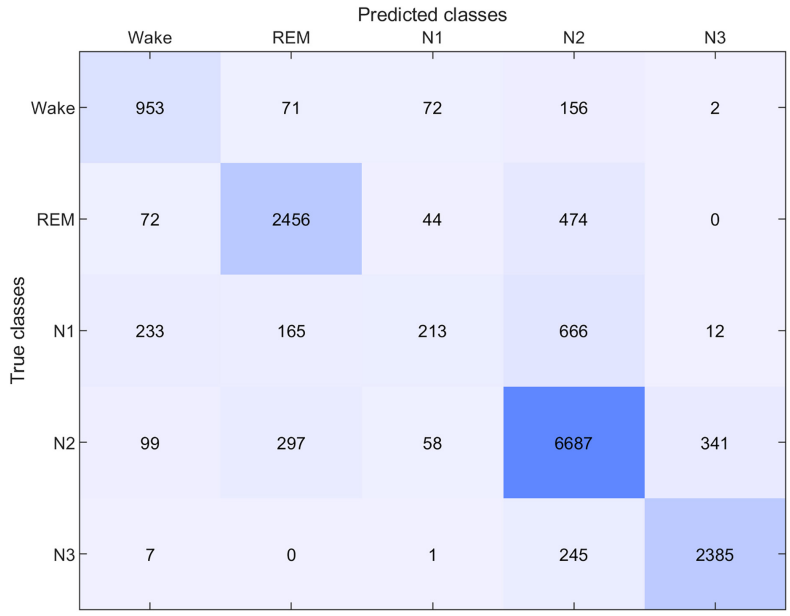


FIGURE 5
Confusion matrix for sleep scoring with the sCGsG cross-validation method.

The average kappa value for the sCsC method was found to be 0.73 which is similar to the value reported in Mikkelsen et al. (2019).

3.4. Effect of equipment, part and subject

We investigated the change of the different sleep metrics with the equipment, part and subject differences. In order to compare the computed sleep metric values between different sets, these values is shown in Figure 8. The distributions of these metrics for sC.A, sC.B, sG.A, and sG.B are presented with different colors. Each circle indicates the value of the related sleep metric for one subject. We observed similar distributions in the data for the different studies and parts. It should be noted that the number of recordings in Part B is considerably larger than in Part A for both studies.

The results of the LMM analysis are presented in Table 2. Each of the models in section "2.5. Automatic sleep scoring" were applied to

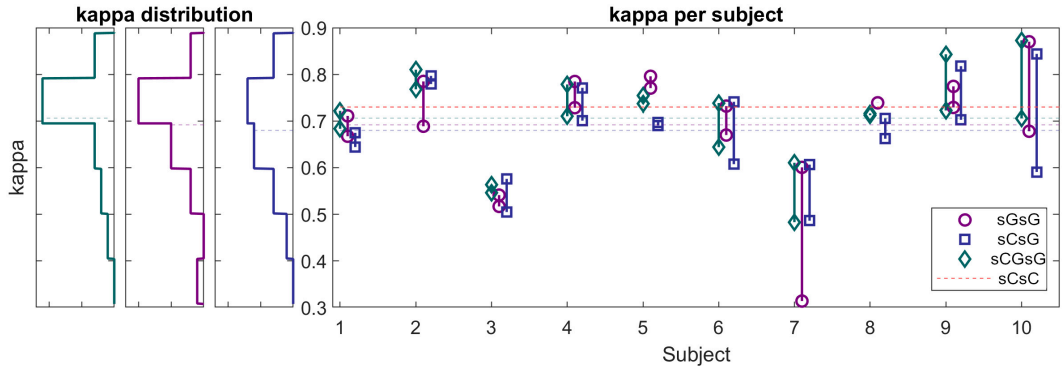


FIGURE 6
(Left panels) Histograms of kappa values for sG recordings (cross-validation scheme: sGsG: purple, sCsG: blue, sCGsG: green). Dashed lines represent the average of each method. For the sGsG and sCGsG schemes, the majority of the recordings yielded a kappa value above 0.7. **(Right panel)** Kappa values for each subject for all three cross-validation schemes. Each point represents one recording.

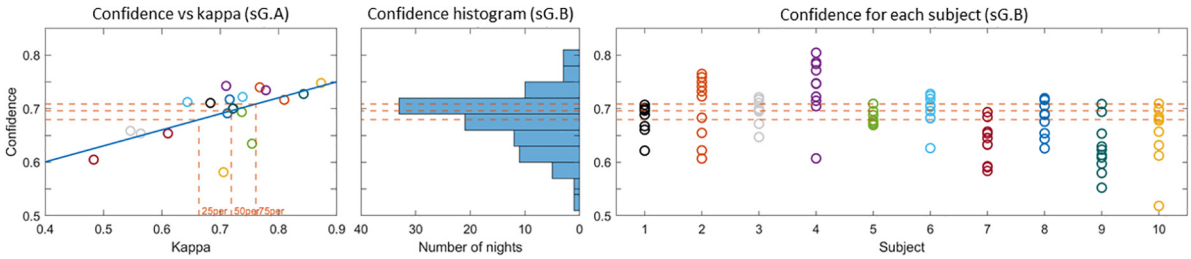


FIGURE 7
(Left) Confidence versus kappa value for the sG.A recordings. The 25th, 50th, and 75th percentiles of the kappa values and the corresponding confidences are shown with dashed lines. **(Center)** Histogram of the confidence values for the sG.B recordings. **(Right)** Confidence values for each recording. The 25th, 50th, and 75th percentile confidence values are presented with dashed lines in all plots.

the sC, sG and sCsG datasets separately. The *p*-value was computed for each fixed effect (subject, part, and equipment).

The Subject parameter had a significant effect on the variation of all metrics for all datasets. Furthermore, the Part effect was only significantly different in the N3fr metric for sCsG dataset. The model coefficient for the fraction of N3 sleep (N3fr) in the combined dataset sCsG was 2.8%. The Part effect was insignificant in all other metric/dataset combinations. The Equipment parameter was also insignificant for all the metrics and datasets.

4. Discussion

This study evaluated a new generic ear-EEG system for sleep monitoring. The system comprised a set of generic earpieces designed to fit most human ears, and the earpieces were connected to a recording device comprising a custom 4-channel "EEG-to-digital converter" ASIC. Both the EEG signal quality and the comfort are essential for a good sleep monitoring device, and the design of the earpieces are important for both these key parameters. The signal quality is intimately related to the electrode-skin interface, and the purpose of the earpiece is to provide a firm and stable electrode-skin connection. The comfort of an earpiece is related to a wide range of factors including mechanical properties of the earpiece material, the number and placement of the electrodes embedded in the earpiece, and the ergonomic design. The earpiece design process was iterative and in each step in the process we evaluated both the comfort and the signal quality. Some of the key experiences obtained in the process, and the consequences for the earpiece design, are summarized here: The earpieces need to be made of a soft and compliant material; in our design the earpieces were made of a silicone material with shore 60. Essential to the ergonomic design is how deep the earpiece goes into the ear canal and how well it fits the anatomy of the ear. In our design the earpiece was relative shallow and was not going more than 5–6 mm into the ear canal.

The signal quality was quantified in terms of the percentage of rejected data epochs after the pre-processing step. Rejection of epochs with poor signal quality is a common practice in sleep monitoring. Poor signal quality can be due to several factors, but the signal quality depends in almost all situations on the electrode-skin contact. Thus, if the signal quality is challenged by e.g., movements, the effect on the signal quality will be further exacerbated by a poor electrode-skin connection. Therefore, the proportion of rejected epochs is a good measure for assessing the quality of the electrode-skin connection–the lower the proportion of rejected epochs the better the electrode-skin interface. In the companion study (sC), using individualized earpieces with six electrodes in each ear and a commercial EEG amplifier, 9% of the epochs were rejected (Mikkelsen et al., 2019). The amount of rejected epochs after the single channel signal construction was 4.3%. In this study (sG), we recorded 20 nights with partial PSG and ear-EEG, as well as 100 nights with only ear-EEG. In these 120 nights of recording, the pre-processing pipeline rejected 10.2% of the epochs, which decreased to 4.4% after single channel signal construction. This value is very similar to our previous rejection

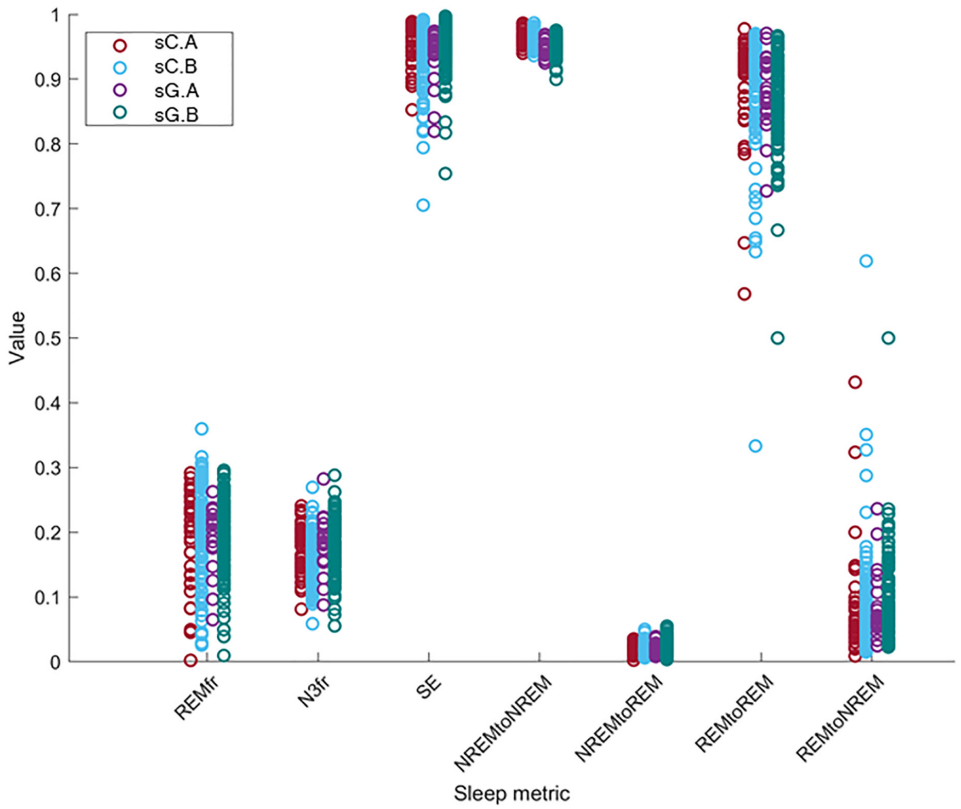


FIGURE 8
The distribution of the value of the sleep metrics (REMfr, N3fr, SE, NREM to NREM, NREM to REM, REM to REM, and REM to NREM) for sC.A, sC.B, sG.A, and sG.B datasets show a large overlap and no clear equipment or part differences.

rate. This demonstrates that changing from custom made to generic earpieces and lowering the number of electrodes did not have a significant effect on the cross-ear single channel signal's quality.

During the study, we identified several factors that caused poor signal quality and occasional rejection of the recording. The main factor was fitting the earpieces. We observed that the subjects were able to easily fit the earpieces after the third or fourth night. Although, the subjects were able to fit the earpieces properly, in some cases, the subject relocated the earpiece during the night without properly fitting it causing the electrodes to lose the connection. In one other case, the subject had excessive movement during the night, which caused the earpiece to fall out of the ear. Otherwise, the subjects were successful in fitting the earpieces. In general, subjects were able to use the device without any problem. Only in few cases, the subjects forgot to charge the device before the recording. We believe that the device can be used easily by a normal user.

Comfort and unobtrusiveness are paramount in long-term sleep recording for at least two reasons: (1) If the sleep monitoring device is not sufficiently comfortable, subjects will not endure it and thus the disadvantages in terms of discomfort will outweigh the benefits of the sleep monitoring. (2) If the monitoring device interferes with the sleep, the sleep information acquired will be biased and thus not provide an accurate impression of the sleep. Therefore, subjects were asked to answer questionnaires after every night to rate their perceived sleep quality, their perceived comfort, the ease of use of the device and whether their sleep had been affected by the earpiece. Based on their answers, we saw that the comfort of the generic earpieces was similar to that of custom-made earpieces from our previous study (Mikkelsen et al., 2019). One of the proposed reasons for this is that it was previously discovered that the depth of the ear canal part of the earpiece was highly correlated with the perceived comfort of the earplug. The generic earpiece was designed to not go further than the second bend of the ear canal, which is the same depth as what the custom earplugs are modeled to. This could explain why the comfort assessments are similar to our previous study. Another important feature in the development of a generic ear-EEG system is the ease-of-use, when asked, most of the subjects reported that the earpieces were easy to mount. In general, subjects preferred the earplugs to the partial PSG setup. The participants responded to comfort and sleep quality questions with a few bad/very bad answers. For most of the questions, the intra subject variation was observed to be higher than the inter subject variation. This shows that the comfort changes more with the recording night rather than the subject.

Another important aspect to consider in this new approach to sleep monitoring is its performance in sleep analysis and scoring. The first part of it, which was mentioned earlier, is the data quality. Poor data quality and missing data is detrimental for sleep scoring. Fortunately, the new earpiece design and the introduction of the customized and application specific amplifier and recording device, did not affect the amount of data rejection. Next, we evaluated the performance of our sleep scoring algorithm on different combinations of this current dataset (sG.A) and a dataset (sC.A) collected in a previous study, with custom earpieces and a commercial amplifier. By combining the datasets, we augment our database and thereby the training and test set of our automatic sleep scoring algorithm. Although the earpiece, amplifier and the number of

TABLE 2 Results from the linear mixed model (LMM) analysis on the sleep metrics.

Metric	Effect	sC	sG	sCsG
REMfr	Subject	$P < 0.01$*	$P < 0.01$*	$P < 0.01$*
	Part	$P = 0.16$	$P = 0.87$	$P = 0.26$
	Equipment	NA	NA	$P = 0.83$
N3fr	Subject	$P < 0.01$*	$P < 0.01$*	$P < 0.01$*
	Part	$P = 0.18$	$P = 0.38$	$P < 0.01$*
	Equipment	NA	NA	$P = 0.46$
SE	Subject	$P < 0.01$*	$P = 0.02$*	$P < 0.01$*
	Part	$P = 0.17$	$P = 0.09$	$P = 0.92$
	Equipment	NA	NA	$P = 0.74$
NREM to NREM	Subject	$P < 0.01$ *	$P < 0.01$*	$P < 0.01$*
	Part	$P = 0.30$	$P = 0.18$	$P = 0.69$
	Equipment	NA	NA	$P = 0.13$
NREM to REM	Subject	$P < 0.01$*	$P < 0.01$*	$P < 0.01$*
	Part	$P = 0.31$	$P = 0.31$	$P = 0.20$
	Equipment	NA	NA	$P = 0.16$
REM to REM	Subject	$P < 0.01$*	$P < 0.01$*	$P < 0.01$*
	Part	$P = 0.28$	$P = 0.43$	$P = 0.49$
	Equipment	NA	NA	$P = 0.43$
REM to NREM	Subject	$P < 0.01$*	$P < 0.01$*	$P < 0.01$*
	Part	$P = 0.11$	$P = 0.39$	$P = 0.23$
	Equipment	NA	NA	$P = 0.49$

The subject effect was significant for all sleep metrics. The part effect was only significant in the combined dataset in the fraction of N3 sleep. The equipment effect was insignificant for all sleep metrics.

*Significant effects are marked.

electrodes were different in those studies, we were able to perform successful automatic sleep scoring by training with one dataset and testing on the other one. First, we observed that training and testing on sG.A yielded a higher performance than training on sC.A and testing on sG.A. This is probably reflecting a combination of two effects: (i) the new setup introduces a certain unique fingerprint in the recordings which is not explained by sC.A. (ii) The characteristic features of sleep in the ear-EEG are largely preserved across the two datasets. Combining the sC.A and sG.A datasets in the training set increased the scoring performance for the sG.A dataset. This combination probably allowed the algorithm to generalize over a larger population of subjects and still learn the unique fingerprints of the new setup. However, combining the datasets did not increase the performance on the old dataset (sC.A). This is likely because sC.A is a sufficiently large dataset for the algorithm to generalize over the population of subjects, therefore adding more training data from sG.A does not add any significant new information. Ultimately, the algorithm continues to have most of its knowledge from sC.A and therefore no increase is observed in the performance on sC.A. The performance of the sCGsG scheme was lower than for the sCGsC scheme. It should be mentioned that sC.A is 4 times larger than sG.A, wherefore train-test on sC.A gives a better generalization. We expect this difference to diminish significantly as our database grows. Subject-wise, we observed that two subjects had considerably lower kappa values compared to the other subjects in all dataset combinations. We suggest that this is caused by different subject specific factors like earplug fitting, sleep characteristics and sleep environment. According to **Figure 5**, the sCGsG method was successful in detecting N2 and N3 stages and less successful in detecting N1 stage. This is similar to what we observed in our previous studies (Mikkelsen et al., 2019; Tabar et al., 2021).

One important aspect of the proposed sleep monitoring system is to provide longitudinal sleep recordings via only ear-EEG recordings. Therefore, it is critical to predict the performance of the sleep scoring algorithm in the part B recordings. While the lack of manual scorings in part B makes it impossible to validate the automatic scorings in the conventional way, we validate them using two analytical methods. The first method is based on the confidence measure derived from the random forest classifier. In a previous study (Mikkelsen et al., 2020), we found a strong correlation between the confidence measure and the kappa value, and this strong relationship was also observed in the current study. This allows us to predict the expected kappa value for part B recordings based on their confidence values. While there is not a one-to-one correspondence between the confidence measure and the kappa value it still provides a good assessment of the performance for the unlabelled data in Part B. As a result, we expect 42% of the Part B recordings to have kappa value above 0.72 which is the median kappa value for part A. Additionally, 58% of the part B recordings were predicted to have kappa value over 0.66 (25th percentile), where 26% of them were predicted to have kappa over 0.76 (75th percentile). While the distribution of the predicted kappa values was highly subject dependent, all the recordings had at least 2 recordings with a kappa above 0.66.

In the second method, we investigated the sleep characteristics of the recordings to see if there was a difference between the estimated sleep characteristics of part A and B for any individual subject. We also aimed to explore whether the new devices (study effect) changed the distribution of the sleep characteristics and ultimately the sleep of subjects. Linear mixed models were constructed for six sleep metrics in which the effects of equipment, part and subject on the sleep metric were investigated. The part effect reflected the effect of the partial PSG setup on the sleep characteristics. We found that the effect of the partial PSG setup was only significant in one sleep metric, namely the fraction NREM stage 3 (N3) sleep for the sCsG dataset. In this case the fraction of N3 sleep increased by 2.8 % from Part A (partial PSG + ear-EEG) to Part B (ear-EEG). The effect of part on all other sleep metrics were insignificant.

Furthermore, we did not find any significant effect of the equipment on any of the sleep metrics, which means that the effect of the new generic earpiece with two electrodes and the new amplifier is negligible. This is in concordance with the fact that we were able to get high classification performance by combining both datasets. The subject effect was significant for all sleep metrics. This shows that changes in the sleep characteristics were related to subject specific idiosyncrasies rather than from which part or which equipment the recordings came from.

These results suggests that generic ear-EEG is a promising method for long-term sleep monitoring. This may have application within e.g., treatment of diseases such as insomnia, chronic pain and many psychiatric disorders, prognostication of recovery after stroke, concussion and traumatic brain injury, and as an early biomarker in neurodegenerative diseases such as Parkinson's disease, Levy Body dementia and Alzheimer's disease.

5. Conclusion

A comfortable and unobtrusive device for long-term sleep monitoring will have great clinical value in diagnosis and treatment of many diseases. We have developed an ear-EEG based sleep monitoring device based on a generic ear-EEG with two recording electrodes in each ear, a proprietary amplifier and an associated automatic sleep scoring algorithm. In this study, we have assessed perceived sleep quality, comfort and ease-of-use and compared the sleep scoring performance against partial PSG scoring. The proposed generic earpiece design was found to be as comfortable as the custom-made design. The EEG signal was recorded from 2 electrodes in each ear with an amplifier made specifically for the current application. The quality of the recorded signal was similar to our previous setup, resulting in successful sleep scoring with an average kappa value equal to 0.71. Automatic sleep scoring was also applied to recordings where no manual scoring was available. We used a confidence measure provided by the sleep stage classifier for assessing the quality of the unlabelled sleep scorings and found that 42% of the part B recordings were estimated to have kappa value above 0.72. Finally, we analyzed the sleep patterns, based on a linear mixed model analysis of seven different sleep metrics, and did not find any statistically significant differences between individualized and generic earpieces, or between PSG-nights and ear-EEG nights. These results suggests that sleep monitoring based on generic ear-EEG devices is a promising alternative to PSG for long-term monitoring of sleep stages.

Data availability statement

The datasets presented in this article are not readily available because the Danish Medicines Agency requires traceability of data in 5 years after completion of the study. To comply with regulations, the data cannot be shared without anonymization of all copies of the dataset. Therefore, data cannot be shared before at earliest at the end of the 5 years period. Until then, the authors are happy to share aggregated statistics presented in this article. Requests to access the datasets should be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by the Central Denmark Region Committees on Biomedical Research Ethics (case no. 1-10-72-13-20). The patients/participants provided their written informed consent to participate in this study.

Author contributions

YT: design of experiment, data collection, ear-EEG data analysis, and manuscript writing. KM: conception of study, design of experiment, and ear-EEG data analysis. NS: analysis of questionnaire data and manuscript writing. SK: experimental setup, design and validation of ear-EEG and PSG-platform. AB, RN, HT, and CH: design and validation of generic ear-EEG earpieces. MH: design of experiment and input to data analysis pipeline. MR: conception of study, design of experiment, and input to data analysis pipeline. MO: supervised PSG scorings and clinical guidance. PK: conception of study, design of experiment, design of ear-EEG platform, data analysis, and manuscript writing. All authors approved the final version of the manuscript and agreed to be accountable for all aspects of the work in ensuring that questions related to the integrity of any part of the work have been appropriately investigated and resolved.

Funding

This research was funded by the Innovation Fund Denmark, grant 7050-00007B.

Conflict of interest

NS, AB, RN, HT, CH, MH, and MR are employed by T&W Engineering A/S, which has a commercial interest in ear-EEG.

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2023.987578/full#supplementary-material

References

Agnew, H. Jr., Webb, W. B., and Williams, R. L. (1966). The first night effect: An Eeg studyof sleep. *Psychophysiology* 2, 263–266. doi: 10.1111/j.1469-8986.1966.tb02650.x

Arnal, P. J., Thorey, V., Ballard, M. E., Hernandez, A. B., Guillot, A., Jourde, H., et al. (2019). The Dreem headband as an alternative to polysomnography for EEG signal acquisition and sleep staging. *Sleep* 43, zsaa097. doi: 10.1101/662734

Berry, R. B., Brooks, R., Gamaldo, C. E., Harding, S. M., Marcus, C., and Vaughn, B. V. (2012). The AASM manual for the scoring of sleep and associated events. *Rules Terminol. Tech. Specific. Darien Illinois Am. Acad. Sleep Med.* 176:2012.

Bleichner, M. G., and Debener, S. (2017). Concealed, unobtrusive ear-centered EEG acquisition: cEEGrids for transparent EEG. *Front. Hum. Neurosci.* 11:163.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Measur.* 20, 37–46. doi: 10.1177/001316446002000104

Gaina, A., Sekine, M., Chen, X., Hamanishi, S., and Kagamimori, S. (2004). Validity of child sleep diary questionnaire among junior high school children. *J. Epidemiol.* 14, 1–4. doi: 10.2188/jea.14.1

Galbiati, A., Verga, L., Giora, E., Zucconi, M., and Ferini-Strambi, L. (2019). The risk of neurodegeneration in REM sleep behavior disorder: A systematic review and meta-analysis of longitudinal studies. *Sleep Med. Rev.* 43, 37–46.

Goverdovsky, V., Looney, D., Kidmose, P., and Mandic, D. P. (2015). In-ear EEG from viscoelastic generic earpieces: Robust and unobtrusive 24/7 monitoring. *IEEE Sensors J.* 16, 271–277. doi: 10.1109/JSEN.2015.2471183

Mikkelsen, K. B., Tabar, Y. R., Toft, H. T., Hemmsen, M. C., Rank, M. L. (2022). "Self-applied ear-EEG for sleep monitoring at home", in *Proceedings of the 2022 44th annual international conference of the IEEE engineering in medicine & biology society (EMBC)*. (Glasgow), 3135–3138. doi: 10.1109/EMBC48229.2022.9871076

Kappel, S. L., Rank, M. L., Toft, H. O., Andersen, M., and Kidmose, P. (2018). Dry-contact electrode ear-EEG. *IEEE Trans. Biomed. Eng.* 66, 150–158. doi: 10.1109/TBME.2018.2835778

Khademi, A., El-Manzalawy, Y., Buxton, O. M., and Honavar, V. (2018). "Toward personalized sleep-wake prediction from actigraphy," in *Proceedings of the 2018 IEEE EMBS international conference on biomedical & health informatics (BHI)*: IEEE), (Piscataway), 414–417. doi: 10.1109/BHI.2018.8333456

Lee, W., Yang, X., Jung, H., Bok, I., Kim, C., Kwon, O., et al. (2018). Anthropometric analysis of 3D ear scans of Koreans and Caucasians for ear product design. *Ergonomics* 61, 1480–1495. doi: 10.1080/00140139.2018.1493150

Looney, D., Park, C., Kidmose, P., Rank, M. L., Ungstrup, M., Rosenkranz, K., et al. (2011). "An in-the-ear platform for recording electroencephalogram," in *Proceedings of the 2011 Annual international conference of the IEEE engineering in medicine and biology society*: IEEE), (Piscataway), 6882–6885. doi: 10.1109/IEMBS.2011.6091733

Mikkelsen, K. B., Kappel, S. L., Mandic, D. P., and Kidmose, P. (2015). EEG recorded from the ear: Characterizing the ear-EEG method. *Front. Neurosci.* 9:438. doi: 10.3389/fnins.2015.00438

Mikkelsen, K. B., Tabar, Y. R., Christensen, C. B., and Kidmose, P. (2021b). EEGs vary less between lab and home locations than they do between people. *Front. Comput. Neurosci.* 15:5. doi: 10.3389/fncom.2021.565244

Mikkelsen, K. B., Phan, H., Rank, M. L., Hemmsen, M. C., De Vos, M., and Kidmose, P. (2021a). Sleep monitoring using ear-centered setups: Investigating the influence from electrode configurations. *IEEE Trans. Biomed. Eng.* 69, 1564–1572. doi: 10.1109/TBME.2021.3116274

Mikkelsen, K. B., Tabar, Y. R., and Kidmose, P. (2020). "Predicting sleep classification performance without labels," in *Proceedings of the 2020 42nd annual international conference of the IEEE engineering in medicine & biology society (EMBC)*: IEEE), (Piscataway), 645–648. doi: 10.1109/EMBC44109.2020.9175743

Mikkelsen, K. B., Tabar, Y. R., Kappel, S. L., Christensen, C. B., Toft, H. O., Hemmsen, M. C., et al. (2019). Accurate whole-night sleep monitoring with dry-contact ear-EEG. *Sci. Rep.* 9, 1–12. doi: 10.1038/s41598-019-53115-3

Mikkelsen, K. B., Villadsen, D. B., Otto, M., and Kidmose, P. (2017). Automatic sleep staging using ear-EEG. *Biomedical engineering online* 16, 1–15. doi: 10.1186/s12938-017-0400-5

Modabber, A., Galster, H., Peters, F., Möhlhenrich, S. C., Kniha, K., Knobe, M., et al. (2018). Three-dimensional analysis of the ear morphology. *Aesthet. Plast. Surg.* 42, 766–773. doi: 10.1007/s00266-017-1027-4

Naismith, S. L., Lewis, S. J., and Rogers, N. L. (2011). Sleep–wake changes and cognition in neurodegenerative disease. *Prog. Brain Res.* 190, 21–52. doi: 10.1016/B978-0-444-53817-8.00002-5

Nakamura, T., Alqurashi, Y. D., Morrell, M. J., and Mandic, D. P. (2019). Hearables: Automatic overnight sleep monitoring with standardized in-ear EEG sensor. *IEEE Trans. Biomed. Eng.* 67, 203–212. doi: 10.1109/TBME.2019.2911423

Sadeh, A. (2015). Iii. Sleep assessment methods. *Monogr. Soc. Res. Child Dev.* 80, 33–48. doi: 10.1111/mono.12143

Stefani, A., and Högl, B. (2020). Sleep in Parkinson's disease. *Neuropsychopharmacology* 45, 121–128. doi: 10.1038/s41386-019-0448-y

Tabar, Y. R., Mikkelsen, K. B., Rank, M. L., Hemmsen, M. C., and Kidmose, P. (2020). "Muscle activity detection during sleep by ear-EEG," in *Proceedings of the 2020 42nd annual international conference of the IEEE engineering in medicine & biology society (EMBC)*: IEEE), (Piscataway), 1007–1010. doi: 10.1109/EMBC44109.2020.9176365

Tabar, Y. R., Mikkelsen, K. B., Rank, M. L., Hemmsen, M. C., Otto, M., and Kidmose, P. (2021). Ear-EEG for sleep assessment: A comparison with actigraphy and PSG. *Sleep Breath.* 25, 1693–1705. doi: 10.1007/s11325-020-02248-1

Toivonen, M., Pääkkönen, R., Savolainen, S., and Lehtomäki, K. (2002). Noise attenuation and proper insertion of earplugs into ear canals. *Ann. Occup. Hygiene* 46, 527–530.

Wilkinson, G., and Rogers, C. (1973). Symbolic description of factorial models for analysis of variance. *J. R. Stat. Soc.* 22, 392–399. doi: 10.2307/2346786

Younes, M., Soiferman, M., Thompson, W., and Giannouli, E. (2017). Performance of a new portable wireless sleep monitor. *J. Clin. Sleep Med.* 13, 245–258. doi: 10.5664/jcsm.6456

Zhou, X., Li, Q., Kilsgaard, S., Moradi, F., Kappel, S. L., and Kidmose, P. (2016). "A wearable ear-EEG recording system based on dry-contact active electrodes," in *Proceedings of the 2016 IEEE symposium on Vlsi Circuits (Vlsi-Circuits)*: IEEE), (Piscataway), 1–2.

Zhu, B., Bronas, U. G., and Fritschi, C. (2018). Sleep assessment in aging adults with type 2 diabetes: Agreement between actigraphy and sleep diaries. *Sleep Med.* 46, 88–94. doi: 10.1016/j.sleep.2018.03.008

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 26 September 2023
DOI 10.3389/fnins.2023.895094

OPEN ACCESS

EDITED BY
Marc Schönwiesner,
Leipzig University, Germany

REVIEWED BY
Grainne McLoughlin,
University of California, San Diego, United States
Jaakko Kauramäki,
University of Helsinki, Finland

*CORRESPONDENCE
Gabrielle Crétot-Richert
✉ gcretot@critias.ca

RECEIVED 12 March 2022
ACCEPTED 12 July 2023
PUBLISHED 26 September 2023

CITATION
Crétot-Richert G, De Vos M, Debener S, Bleichner MG and Voix J (2023) Assessing focus through ear-EEG: a comparative study between conventional cap EEG and mobile in- and around-the-ear EEG systems.
Front. Neurosci. 17:895094.
doi: 10.3389/fnins.2023.895094

Assessing focus through ear-EEG: a comparative study between conventional cap EEG and mobile in- and around-the-ear EEG systems

Gabrielle Crétot-Richert[1]*, Maarten De Vos[2], Stefan Debener[3,4], Martin G. Bleichner[4,5] and Jérémie Voix[1]

[1]École de technologie supérieure (ÉTS), Université du Québec, Montréal, QC, Canada, [2]Stadius, Department of Electrical Engineering, Faculty of Engineering Sciences & Department of Development and Regeneration, Faculty of Medicine, KU Leuven, Leuven, Belgium, [3]Neuropsychology Lab, Department of Psychology, University of Oldenburg, Oldenburg, Germany, [4]Research Center for Neurosensory Science, University of Oldenburg, Oldenburg, Germany, [5]Neurophysiology of Everyday Life Group, Department of Psychology, University of Oldenburg, Oldenburg, Germany

Introduction: As our attention is becoming a commodity that an ever-increasing number of applications are competing for, investing in modern day tools and devices that can detect our mental states and protect them from outside interruptions holds great value. Mental fatigue and distractions are impacting our ability to focus and can cause workplace injuries. Electroencephalography (EEG) may reflect concentration, and if EEG equipment became wearable and inconspicuous, innovative brain-computer interfaces (BCI) could be developed to monitor mental load in daily life situations. The purpose of this study is to investigate the potential of EEG recorded inside and around the human ear to determine levels of attention and focus.

Methods: In this study, mobile and wireless ear-EEG were concurrently recorded with conventional EEG (cap) systems to collect data during tasks related to focus: an N-back task to assess working memory and a mental arithmetic task to assess cognitive workload. The power spectral density (PSD) of the EEG signal was analyzed to isolate consistent differences between mental load conditions and classify epochs using step-wise linear discriminant analysis (swLDA).

Results and discussion: Results revealed that spectral features differed statistically between levels of cognitive load for both tasks. Classification algorithms were tested on spectral features from twelve and two selected channels, for the cap and the ear-EEG. A two-channel ear-EEG model evaluated the performance of two dry in-ear electrodes specifically. Single-trial classification for both tasks revealed above chance-level accuracies for all subjects, with mean accuracies of: 96% (cap-EEG) and 95% (ear-EEG) for the twelve-channel models, 76% (cap-EEG) and 74% (in-ear-EEG) for the two-channel model for the N-back task; and 82% (cap-EEG) and 85% (ear-EEG) for the twelve-channel, 70% (cap-EEG) and 69% (in-ear-EEG) for the two-channel model for the arithmetic task. These results suggest that neural oscillations recorded with ear-EEG can be used to reliably differentiate between levels of cognitive workload and working memory, in particular when multi-channel recordings are available, and could, in the near future, be integrated into wearable devices.

KEYWORDS

EEG, brain computer interface, cEEGrid, attention, cognitive workload, working memory, machine learning, ear-EEG

1. Introduction

The push to develop and democratize useful neurotechnology tools and devices has been fueled by both the academic and industrial sectors. Over the last twenty years, neuroscience as a field of research has grown remarkably, generating major funding and institutional support (Zivkovic, 2015). The private sector has also increased investments and a handful of companies are now considered leaders of innovation in the neurotech landscape, not without growing ethical concerns and public scrutiny (Jarchum, 2019; Pfotenhauer et al., 2021; Wexler, 2021). In this context, new applications are emerging for medical as well as consumer interests and with these come new challenges, specifically, the need for improved mobility (Debener et al., 2012; Gramann et al., 2014). Transferring knowledge and technology from a laboratory environment to real-world applications is both necessary and far from trivial. Electroencephalography (EEG) is a proven technique to record brain-electrical activity with the important advantage of high temporal resolution. Non-invasive EEG uses electrodes placed on the scalp to capture electrical potentials emitted by large groups of neurons firing synchronously. EEG is a safe, low-cost and low energy technology compared to other brain imaging techniques, such as magnetic resonance imaging (MRI). This has made EEG particularly appealing for brain-computer interfaces (BCI). BCIs allow humans to control electronic devices, such as computers or prosthetics (Hochberg et al., 2006), using their thoughts. Earlier uses of such devices mainly focused on communication and control for individuals suffering from different forms of paralysis (Birbaumer et al., 1999; Wolpaw et al., 2002; Vansteensel et al., 2016). However, other, more quotidian BCI applications are emerging (Zander and Kothe, 2011). For instance, BCI applications could monitor the mental state of a person in view of workplace security (Müller et al., 2008; Aricò et al., 2016; Mijović et al., 2017) and productivity, by preventing interruptions that could be detrimental to task completion and quality (Jenkins et al., 2016). If we could reliably decode when a person is focused on a task, we could protect that state of "flow" by limiting both visual and auditory distractions. This study intends to do the former using non-invasive, discrete technology that will not limit movement.

The robustness of BCI devices is improving thanks to recent progress in signal processing, machine learning, electrode technology and open-source software tools (Delorme and Makeig, 2004; Popescu et al., 2007; Blankertz et al., 2011; Gramfort et al., 2014). However, several challenges still need to be addressed. For instance, the recording equipment usually involves caps or headbands, making it unsuitable in social settings while the electrodes are usually wire-connected to bulky amplifiers making the devices cumbersome and stationary. Thus, recent EEG technological developments have been focusing on lighter equipment with fewer electrodes and greater mobility, more adapted for exploration of cognitive processes in realistic environments and situations (Casson et al., 2010; Chi et al., 2011; Debener et al., 2015; Goverdovsky et al., 2017). Wireless EEG amplifiers are becoming available and their capability to record reliable EEG data has been scientifically proven (Debener et al., 2012; De Vos et al., 2014b; Lin et al., 2014).

Ear-EEG refers to the recording of brain-electrical activity from electrodes placed in or near the ear, as opposed to traditional scalp EEG, for which electrodes are commonly placed in concentric circles on caps which cover the entire scalp. The main appeal of ear-EEG for the proposed mental focus application is that it can be inconspicuous to wear in public settings. Also, the area around the ear and inside the ear canal is usually hairless, an advantage for electrode-skin contact with or without conductive gel. Furthermore, the irregular geometry of the ear canal would allow fitted earpieces to keep the device firmly in place, thereby potentially reducing artifacts during natural movement. The first proof-of-concept for ear-EEG was published in 2011 (Looney et al., 2011) and since then, multiple research groups have demonstrated its potential, with electrodes placed around the ear, in the concha and inside the ear canal (Debener et al., 2012; Kidmose et al., 2012; Bleichner et al., 2015). Ear-EEG has already proven its capacity to reliably record auditory attention (Bleichner et al., 2016; Mirkovic et al., 2016), event-related potential (ERP) components such as the P300 (Debener et al., 2015; Pacharra et al., 2017) and factors like fatigue and sleep quality (Looney et al., 2014; Mikkelsen et al., 2018; Sterr et al., 2018). Some studies have considered the potential of ear-EEG to record alpha frequencies (rhythmical activation of the brain oscillating between 8 and 12 Hz) (Debener et al., 2015; Mikkelsen et al., 2015), but ear-EEG research has generally focused on the analysis of time-locked events. ERPs are induced potentials that are seen in response to sensory, cognitive or motor events. They combine positive and negative amplitude peaks occurring over specific time windows after stimulus onset, some are large in amplitude and identifiable even at the single trial level (e.g., P300 or N100). Another type of brain activity, which has received perhaps less attention in ear-EEG research, is continuous EEG and its oscillatory behaviors (Buzsaki, 2011). Neural oscillations are defined as power spectral densities over specific frequency ranges, delta (1–3 Hz), theta (4–7 Hz), alpha (8–12 Hz), beta (13–30 Hz) and gamma (> 30 Hz). They have been found to reflect mental states such as memory (Klimesch et al., 1996), attention (Foxe and Snyder, 2011), engagement (Berka et al., 2007) and higher thinking and reasoning processes (Palaniappan, 2006).

This study aims to investigate the EEG power spectrum extracted from signals recorded in and around the ear and to evaluate the potential to use these signals to monitor mental states. We compared conventional cap-EEG (EEG recorded over the entire head) to new devices developed for mobile ear-EEG to assess how the signals' amplitudes and power spectrum differ from one recording system to the other. The goal is to discriminate between states of high and of low focus using classification algorithms (Müller et al., 2008).

In this study, focus is associated to two main concepts: working memory and cognitive workload as seen in (Klimesch et al., 1998; Oken et al., 2006; Fougnie, 2008). *Working memory* refers to the process of using short-term memory to make immediate and conscious decisions regarding a perceptual or linguistic task. A popular paradigm used to investigate working memory is the N-back task as presented in Brouwer et al. (2012). Different working memory loads can be studied by lengthening or shortening the duration of a sequence of numbers or letters a subject is asked to remember. Working memory load is known to inversely correlate

with alpha power (Owen et al., 2005; Berka et al., 2007; Regenbogen et al., 2012). Additionally, brain oscillations from theta to gamma have been found to vary according to the working memory levels (Herrmann et al., 2004; Pesonen et al., 2007). The N-back task elicits a P300 ERP, a well-known ERP component, which varies in amplitude according to task relevance and working memory (Watter et al., 2001). *Cognitive workload* relates to the difficulty and effort required to perform a task. It has strongly been associated with changes in alpha power according to the level of difficulty of the task (Anderson and Sijercic, 1996; Keil et al., 2006; Foxe and Snyder, 2011; Magosso et al., 2019). It has been reported that other frequency bands, such as delta (Harmony et al., 1996), theta (Scheeringa et al., 2008), and gamma (Shibata et al., 1999; Landau et al., 2007) also carry information relative to cognitive workload. To assess the ear-EEG's capability to detect changes in power spectral density (PSD) relevant to cognitive workload levels, an arithmetic task was replicated from established high-density cap-EEG studies (Yu et al., 2009; Rebsamen et al., 2011).

2. Materials and methods

2.1. Subjects

Fifteen subjects (nine females, six males, mean age 27.1 years, 14 right-handed) were initially recruited for this study to participate in both tasks although some participants only took part in one of the two. Each gave written informed consent and none reported any neurological or psychiatric disorders. Recruitment and procedures for this study were performed in accordance with the Ethics Committee of Oldenburg University. Subjects were remunerated at minimum wage to take part in the study.

2.2. Test paradigm

Two tasks were chosen to study different characteristics and common features of concentration: an N-back task to study working memory and an arithmetic task to study cognitive workload. For both tasks, subjects were seated in front of a screen where visual stimuli, letters and numbers were displayed using the Presentation® software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com), which also generated event markers. The stimuli were displayed in a black font on a light gray background. For each task, subjects were first given a tutorial round to ensure the instructions were understood. Task order was balanced across subjects.

The N-back task was conducted according to the following study's procedure (Brouwer et al., 2012). Subjects memorized a series of consonants and for each new letter that appeared, they had to decide if it was a target or a non-target letter using a two-button handheld box. Button assignment for target and non-target, left or right, was balanced across subjects. The attributes of the target letter depended on the level of difficulty of the task, of which there were three. The first level or 0-back, required no working memory effort as the subjects did not have to pay attention to the sequence of letters. They only had to identify the letter 'X' as a target and any other letter as a non-target. For the 1-back, subjects needed to remember the past letter shown. If the new stimulus was the same as the previous one, one back, then that letter was a target. And finally, for the 2-back, the most demanding level in terms of working memory, the target letters were those shown two letters before, requiring the subjects to constantly update the last two letters in their head and compare it with the new one. The letters were displayed on the screen for 500 ms with a 2000 ms inter-stimulus interval during which a fixation cross was displayed at the center of the screen.

The duration of the N-back task was about 45 min. It was divided into four sessions of six two-minute blocks each. The three levels of difficulty were repeated twice during one session. Forty-eight letters were shown during each block, 33% of which were targets. The levels were given in a pseudorandom order, different for each subject. Each level was presented once before being repeated and the same level was never given twice in a row. Between each block, the subjects were shown their success rate on the screen followed by a fixation cross for 20 seconds. Subjects initiated the next block and between sessions, the experimenters would briefly interact with the subjects before initiating the next session.

The arithmetic task consisted of sums and inequalities, a procedure adapted from Rebsamen et al. (2011). Subjects were shown an addition, asked to calculate it and keep the result in mind. A new number was then displayed which could be greater, equal to or lower than the result, each with a 33% probability. Subjects were asked to compare their answer with this number, using a handheld 3-button box. The buttons were marked with the signs "<", "=" and ">". There was no time limit to solve a problem. The task featured five levels of difficulty; level 1: addition of two single-digit numbers, level 2: addition of a single and a double-digit number, level 3: addition of two double-digit numbers, level 4: one double and one triple-digit number and finally, level 5: an addition between two triple-digit numbers.

The arithmetic task took 35 min to complete. It was divided into two sessions, each consisting of fifteen one-minute blocks and the five levels of difficulty were repeated three times. During a block, only additions of the same level were featured. Subjects completed as many additions as they could in the one-minute time frame. The levels of difficulty were given in a pseudorandom order, different for each subject. Each level was presented once before being repeated, and the same level was never given twice in a row. Between each block, subjects were shown their success rate on the screen. A 20-s rest period followed each block, during which a fixation cross was displayed on the screen. Between sessions, experimenters would interact with the subjects until they felt ready to initiate the next session.

2.3. Data acquisition

EEG data was recorded using two recording systems concurrently: an ear-EEG mobile recording system and a high-density cap-EEG stationary recording system as described in Bleichner et al. (2016). The ear-EEG equipment consisted of a SMARTING 24-channel wireless EEG amplifier (mBrainTrain,

Belgrade, Serbia) with a modified connector to two cEEGrids–concealed, around the ear self-adhesive arrays of 10 flex-printed Ag/AgCl electrodes (Debener et al., 2015) - and two TIPtrode™ - in-ear foam inserts wrapped in gold foil (Bauch and Olsen, 1990) as seen in Figure 1. Ear-EEG channels, cEEGrids and TIPtrodes™, were referenced to a channel on the cEEGrid placed around the subject's right ear, electrode R4b, while electrode R4a served as the analog ground. Channels from the right ear, cEEGrids and TIPtrodes (excluding R4a and R4b), were re-referenced during signal pre-processing to channel L4b ensuring that each ear-EEG channel was referenced to a channel on the contralateral ear. Data from this system was transmitted wirelessly via Bluetooth to the recording computer.

The cap-EEG system consisted of a research-grade EEG amplifier (Brainamp, Brainproducts GmbH, Herrsching, Germany) connected to a 96-channel Ag/AgCl EEG cap (EasyCap, Hersching, Germany) with equidistant electrodes. Twelve electrodes were not connected, that is, six over each ear, as they would have overlapped with the cEEGrids placed underneath the cap. This gave a total of 84 effective channels. The ground was placed at a central fronto-polar electrode and the reference electrode was placed at the nose-tip. This system had two dedicated eye electrodes placed under each of the participant's eyes (E29 and E30). The cap-EEG data was transmitted through fiber optic cable to the same recording computer as the one used for the ear-EEG. Both recording systems' data streams were recorded at a 500 Hz sampling rate along with a third data stream consisting of the event markers generated by the Presentation® software (Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). As shown in Figure 2, they were combined into a single file, xdf format, using the Lab Recorder program from the open-source Lab Streaming Layer (LSL) data acquisition and synchronization software (Kothe, 2014).

Prior to recording, subjects were asked to wash and dry their hair. They were also given cotton swabs to clean earwax from the outer ear canal. Alcohol wipes and abrasive gel were used to prep the skin around the ear. Electrolytic gel (Abralyt HiCl, Easycap GmbH, Germany) was applied to the cEEGrid electrodes, which were then placed around the subjects' ears. Electrode impedance was adjusted to under 20 kΩ for each ear electrode. Next, the subject was fitted with the electrode head cap. The six electrode rings over each ear were taped off, all others were covered with electrolytic gel and individual electrode impedance was set to under 5 kΩ for all cap electrodes. The subjects were seated facing a screen at a distance of 1.3 meters, in a sound-proof room. Lastly, the two dry TIPtrodes™ were rolled tight and inserted into the ear canal where they were given time to expand and come in close contact with the skin. Metallic clamps were attached to the gold foil of the TIPtrodes™ and connected to a SMARTING adaptor along with the cEEGrids leads. The SMARTING amplifier was secured to the shoulder of the subject.

2.4. Data analysis

Data was analyzed offline using MATLAB (MathWorks Inc., Natick, USA) and EEGLAB, version 13.6.5b (Delorme and Makeig, 2004). Fourteen subjects took part in the N-back task and all were included in this study. Fifteen subjects were recruited for the arithmetic task, but five of these were not considered because of a stimuli marker experimental error that failed to associate EEG data blocks to their level of difficulty. EEG data was band-pass filtered between 0.5 Hz and 100 Hz and resampled at 256 Hz. The data sets were then epoched for each task. For the N-back task, epochs started 500 ms before a new stimulus was presented and ended 1500 ms after. Epochs were baseline corrected at 200 ms before stimulus presentation. This resulted in 48 epochs per block, 384 per level and 1,152 per subject. For the arithmetic task, epochs were not linked to stimuli presentation since the number of problems and the time lapse between problems varied across levels and subjects. Therefore, data from the arithmetic task was epoched at regular intervals. Regularly-spaced dummy event markers were added during pre-processing and used to epoch the EEG data in 2-second windows with a one-second overlap between windows. This produced 58 epochs per block, 348 epochs per level and 1,740 epochs per subject). Epochs from both tasks were inspected visually to select artefact-dominated trials and remove them. Less than 5% of trials were removed for each individual subject data set. An Independent Component Analysis (ICA) was then performed on all available channels for each recording equipment (84 for the cap-EEG data and 20 for the ear-EEG). The eye blink component or components were removed using topographical distribution when available and time courses of the independent component activation for both systems. Between one and three components was removed per subject per task as the eye blink component(s) for the cap-EEG data while for the ear-EEG data, components removed varied between one and two components per subject per task, with an exception of one participant for which no component were removed. This participant's ear-EEG data exhibited in fact very few blink related activity. For both the cap and the ear EEG data, the blink removal procedure was validated by inspecting superimposed single trial time courses before and after component removal using EEGLAB tools to ensure that the blinks had been removed from the signal without affecting other dynamics of the continuous EEG signal. Other components identified as noise, notably an important sinusoidal component which affected some cap-EEG electrodes located at the back of the ears and both in-ear-EEG electrodes, were removed using the ICA method as well. This noise might have been related to interference between the two recording systems and equipment.

Data sets were then processed for a spectral-domain analysis. The Power Spectral Density (PSD) was calculated for each epoch using a periodogram power spectral density estimate under a Hanning window. The results were interpreted for a frequency range from 1 to 100 Hz in 1 Hz interval bins and as band power for seven neural oscillation groups: delta (1–3 Hz), theta (4–7 Hz), alpha (8–12 Hz), a low beta band (13–19 Hz) and a high beta band (20–30 Hz), a low gamma band (30–50 Hz) and a high gamma band (50–100 Hz). This splitting of the last two frequency bands was based on the results of an analysis, not shown here, which came to a compromise between showcasing significant differences in PSD within the bands while limiting the amount of splits in the bands when the differences were not as statistically significant.

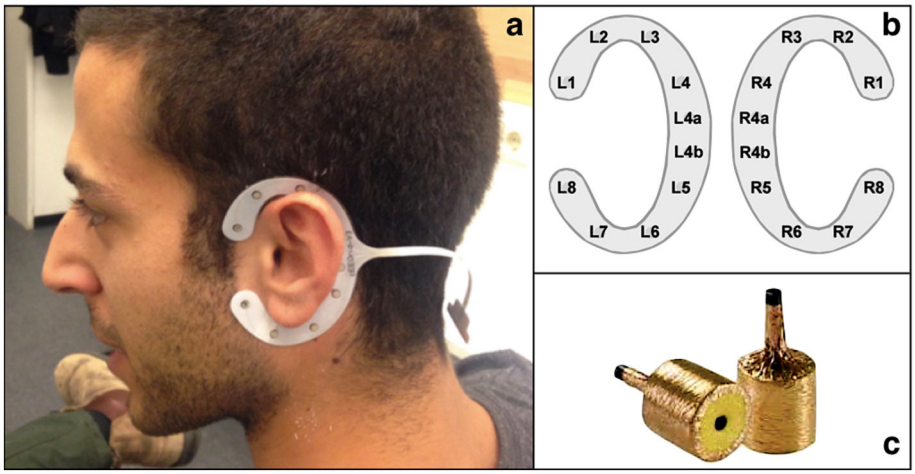


FIGURE 1
Ear-EEG recording devices: **(a, b)** show the cEEGrid; **(c)** shows the TIPtrodes™.

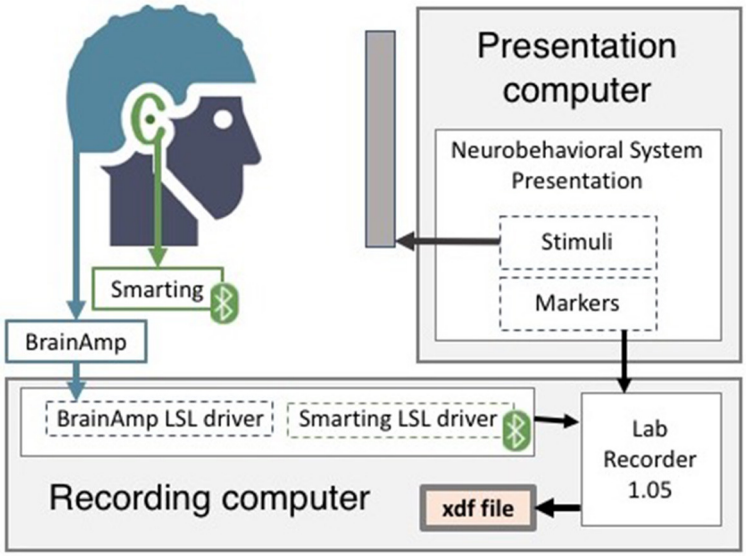


FIGURE 2
EEG devices setup and recording systems.

A time-domain analysis was performed for the N-back task, for which the data was resampled at 100 Hz and a low-pass filter was applied at 20 Hz to analyze the ERP response from 0 to 1,000 ms.

For both these analyses, spectral and temporal, all channels were studied and two electrodes from each recording system were selected to highlight grand average and distribution results: a centrally-located electrode, Pz for cap-EEG, L3 for ear-EEG; and another electrode which was relevant to compare in-ear EEG to cap-EEG specifically. These special electrodes were: one of the in-ear channels, E1 and E2, the two TIPtrodes™ electrodes used dry inside the ear canal; and the cap-EEG channel chosen for comparison consisted of an occipital channel, O1 or O2, chosen after multiple considerations of other electrodes (midline, central and frontal pairs). The analysis (and later classification) results were found to be best at these locations which was to be anticipated considering the visual nature of both tasks' experimental stimuli. Only one in-ear channel and one occipital channel were shown in the figures since their results looked similar enough that no additional information could be gathered from representing both.

A spatial-domain analysis was performed using all 84 cap-EEG channels available. Topographical distributions for the above-mentioned seven frequency band powers was mapped for different mental load conditions. The difference between the most and least demanding conditions and the relative significance of these differences were also represented spatially.

Lastly, statistical analyses of the differences between mental load conditions was done through permutation testing (Cohen, 2014) for twelve electrodes from both recording systems which

covered the recorded channels as well as possible. The cap-EEG's channels provided a good spatial distribution over the head and were based on De Vos et al. (2014b). Since the cap-EEG electrode layout used in this study differed from the cited article's layout, a correspondence was established between the 10–20 system electrode labels and the 96 equidistant electrode cap labels using the shortest Euclidean distances between electrode positions. The cap-EEG electrodes chosen for this study are: E1-Cz, E2-Fz, E5-Pz, E8-Fpz, E12-O2, E13-O1, E45-F4, E38-C4, E49-P4, E53-P3, E41-C3, E57-F3. The ear-EEG's channels which were found to be representative of the entire electrode set and overlapped with the results of prior authors (Bleichner et al., 2016). They are: L1, L2, L3, L6, L7, R1, R2, R3, R6, R7 for the cEEGrid plus the right and left in-ear TIPtrode^{TM} electrodes referred to as E1 and E2. This non-parametric test was chosen because PSD distributions from all subjects were largely non-uniform. Additionally, the statistical results of a permutation test, a z-score, is a signed indicator which gives information on the strength of the statistical test but also on the sign of the difference between conditions. It was shown that the sign of the differences - which condition yields a higher PSD or amplitude and which is lesser - changes between mental load conditions according to individual subject, frequency, time and spatial location. Each permutation test was done 5,000 times to yield z-scores for each comparison and significance threshold was placed at a p-value of 0.01, equivalent to a z-score of ± 2.326. The results of these comparisons are featured in the analysis of the EEG data for both tasks. Statistical heatmaps were created from them for both spectral and temporal domain information. To account for the important number of tests performed to produce each heatmap (over a thousand), multiple comparison correction was applied to the statistical values using the False Discovery Rate (FDR) method (Benjamini and Hochberg, 1995; Yekutieli and Benjamini, 1999).

2.5. Feature extraction and classification

Classification of the data epochs was performed on a subset of the total number of channels available (twelve and two) for each recording system to be consistent across both recording systems and more representative of concealed ear-EEG device constraints. Narrowing the number of channels considered was also important for dimensionality concerns associated with machine learning algorithms. Multiple models were generated using one classifier type, a step-wise linear discriminant analysis (swLDA), each model was defined by the features of the EEG signal it used. Sets of features were generated in order to compare their performance and assess the influence of factors such as the number of channels available, the PSD resolution, the cut-off frequency and the domain of the features (spectral vs temporal). The same channels as for the statistical heatmaps were used to create the twelve-channel feature sets for both recording equipment and both task. The channels used to compare in-ear to cap-EEG specifically were used to generate the two-channel features sets: E1 and E2 for the ear-EEG, O1 and O2 for the cap-EEG.

PSD features for each channel were extracted from the periodogram of the epochs generated for each task and focus level. The main PSD feature set considered was composed of PSD estimates from 1 to 100 Hz with a 1 Hz resolution also referred to as the 1 Hz frequency bin model. It generated 100 initial features per channel. Each model was trained for both cap and ear-EEG, for twelve channels (1200 initial features) and for two channels (200 initial features). The PSD values for each selected channel were concatenated.

To study the influence of PSD resolution on classifier performance, models with different PSD resolution or bin sizes were considered, two additional frequency bin models and 3 frequency band models. The highest resolution model was composed of 0.5 Hz bins, yielding 200 initial features per channels. A 5 Hz or less PSD model was evaluated, it was composed of a delta (1-3 Hz), theta (4-7 Hz) and alpha (8-12 Hz) neural oscillation bands since their frequency ranges were equal to or lower than 5 Hz; the remaining interval (13 to 100 Hz) was divided into 5 Hz-frequency bins resulting in 21 features per channel. The twelve-band model (or 10 Hz model) started the same with frequency bands delta to alpha, it split the beta frequency range in two (much like the seven frequency band model described earlier) for a low beta (13 to 19 Hz) and a high beta (20 to 30 Hz), the gamma band was then equally split in 10 Hz frequency bins from 30 to 100 Hz for a total of 12 features (or bands) per channel. Refer to the previous section for the seven-band model. The five-band model, the lowest resolution model, consisted simply of the five main neural oscillations: delta, theta, alpha, beta (13-30 Hz) and gamma (30 Hz to cut-off frequency). These PSD feature sets were compared to the 1 Hz frequency bins model and to each other. The last PSD feature model were created to study the influence of cut-off frequency. The gamma band cut-off frequency appears inconsistent in the EEG literature: Buzsaki (2011) uses both 80 Hz and 100 Hz as a gamma band cut-off at multiple occasions in his book on brain oscillations. Other researchers suggest that gamma band activity could extend to 200 Hz (Uhlhaas et al., 2011), while representations of EEG grand average PSD often stop at 45 Hz. Hence, feature sets were extracted with a 100, 80, and 45 Hz cut-off frequency.

Because the N-back task had temporal information from the ERP response available as well as PSDs, additional models were considered for this task inspired by the following study (Brouwer et al., 2012). One model consisted of temporal features from the time-locked epochs: the amplitude of the signal from 0 to 1,000 ms taken every 10 ms, (100 initial features per channel). Lastly, a mixed set of features, dubbed the "fusion" features by Brouwer et al., incorporated the 100 temporal features and 1 Hz interval PSD features for a total of 200 initial features per channel.

The choice of a swLDA classification algorithm was made since it greatly reduced the number of features considered, thereby reducing the risk of over-fitting (Blankertz et al., 2011), important to address considering the high number of features some models have compared to the number of trials. This technique has proven very powerful for EEG single-trial classification (Krusienski et al., 2006). The swLDA was implemented according to specifications from De Vos et al. (2014b). Feature selection was done sequentially and features were only included if they improved class discrimination statistically ($p_{in} < 0.1$). For each new feature included, the procedure re-analyzed the current feature pool and removed any feature that had become redundant ($p_{out} > 0.15$).

All classifier models in this study were binary. For the arithmetic task, conditions were defined as high and low cognitive workload state. Data from levels 1 and 2 were combined to form the low cognitive workload condition, and levels 4 and 5 were combined to form the high cognitive workload condition. Level 3 was left out to maintain a balanced set of trials per class. This produced a set of over 1250 samples across 2 classes for low vs. high cognitive workload conditions. For the N-back data, high and low working memory conditions were associated with the 2-back and 0-back. This data set held over 700 samples over 2 classes for the high vs. low condition.

The performance of each model was obtained using a ten-fold cross validation test. Before training, the models were divided into 10 sub-groups. Each trial was given a number from 1 to 10, ensuring the same partition of trials or samples was used to evaluate the performance of all the feature sets. The classifiers were trained on all the subgroups except for one, the testing set, which would be used to evaluate the classification accuracy of the trained model. The model then classifies each trial from the testing set using the knowledge it gathered during the training phase. Classification accuracy corresponds to the percentage of trials in the testing set correctly classified as low or high focus by the model. The process was repeated so that each subgroup served as the testing set, yielding ten accuracy results for each model and subject. Reported subject accuracy was the average of these 10 results.

The threshold for above chance-level classification accuracy of single subject data was set using the method in Combrisson and Jerbi (2015), which considers sample size to determine chance-level. Seeing as there were more than 500 samples for each task, in the high vs. low and pairwise binary classification, and that these were 2-class determinations, the chance-level was set at 57% (for a p-value of 0.001).

To compare statistical performance of cap and ear-EEG classifiers and assess the influence of different feature extraction and selection methods, a paired t-test was used on the performance metrics. Each model had 10 accuracy results per subject, a p-value of 0.001 was considered significant, some average p-values were reported when relevant.

3. Results

3.1. EEG data

3.1.1. N-back task

The N-back task's data epochs were locked to the presentation of a stimuli. This made it possible to study PSD and ERP amplitudes to assess the fluctuations in EEG activity between low and high working memory conditions. The difference between these states was assessed on group data from all subjects and on an individual subject basis to create statistical heatmaps.

Figure 3 shows the grand average PSD for two cap and two ear-EEG channels for the three levels of the N-back task on a logarithmic scale. The grand average PSD shows three distinct conditions visible with both recording systems, ear and cap, particularly around the alpha peak (8–12 Hz) for all four channels represented. The amplitudes of the alpha peaks are higher for the cap-EEG channels, Pz and O1, compared to the ear-EEG channels. Conditions with higher working memory load exhibited visibly weaker alpha peaks over both recording systems. It also appears that this order, with the least demanding condition on top, the moderate condition in the middle and the more demanding condition at the bottom seems to be the same over the theta range (4–7 Hz) and the lower part of the beta range (below 20 Hz). This holds true for virtually all channels available over both recording systems with a few exceptions where PSD levels were less distinctive. Around 20 Hz and to the end of the frequency range considered, the order of the plot lines gets inverted with the more demanding condition now generating higher PSD values over the high beta range (from 20 to 30 Hz) and gamma range (30 Hz and above).

In Figure 4, the normalized PSD for different neural oscillations for all subjects is represented with boxplots. The range is divided into seven frequency bands: delta, theta, alpha, two beta bands, low beta (beta 1 for 20-30 Hz) and high beta (beta 2 for 20–30 Hz) and two gamma bands, low gamma (gamma 1 30-50 Hz) and high gamma (gamma 2 50–100 Hz). The beta band was split to coincide with the shift seen in the grand averages of Figure 3. The gamma band mid-point between low and high was chosen after it was determined that more splits or different frequency splits than 50 Hz did not have a significant impact on representation or statistical results. In order to visualize power bands from all the frequency ranges in one figure, band power values were normalized at the individual subject level (divided by the 90th percentile for this band power). Normalized values from all subjects were then pooled together and plotted. The data sets have a substantial amount of outliers and many with extreme values. To reduce the number of outliers and improve readability, the interquartile range used for the whiskers of the boxplots is extended to 2.5. Outliers with extreme values are compressed if they are above 2.5 normalized PSD. A dotted line is placed at this upper limit and outlier markers are plotted evenly in the region above it. Z-scores obtained from permutation testing over trials from all subjects are given over each boxplot. The higher the absolute value of the z-score, the more significant the difference between the two mental load conditions. The sign of the z-score indicates the direction of this difference (i.e. positive values indicate the more demanding condition, the 2-back in this case, yields higher PSD values than the less demanding condition, the 0-back).

From Figure 4, it can be noted that most of the distributions are different and highly significantly from their z-scores, for all channels considered. Though the boxplots themselves can sometimes be harder to interpret, particularly with median values, the same trends as with the grand averages (Figure 3) can still be observed over both recording systems. Box outlines, whisker limits and outlier values hold noticeable differentiating information between conditions. That is a tendency for smaller PSD values as mental load increases for the lower frequency bands (under 30 Hz). All channels represented had negative Z-scores for the delta, alpha and beta 1 bands. The alpha bands are the most distinguishable, yielding the highest negative z-scores for every channel considered. While from the beta 2 band onward the PSD is now higher with increasing mental load. All gamma bands z-scores are negative and highly significant and although beta 2's z-score is positive for

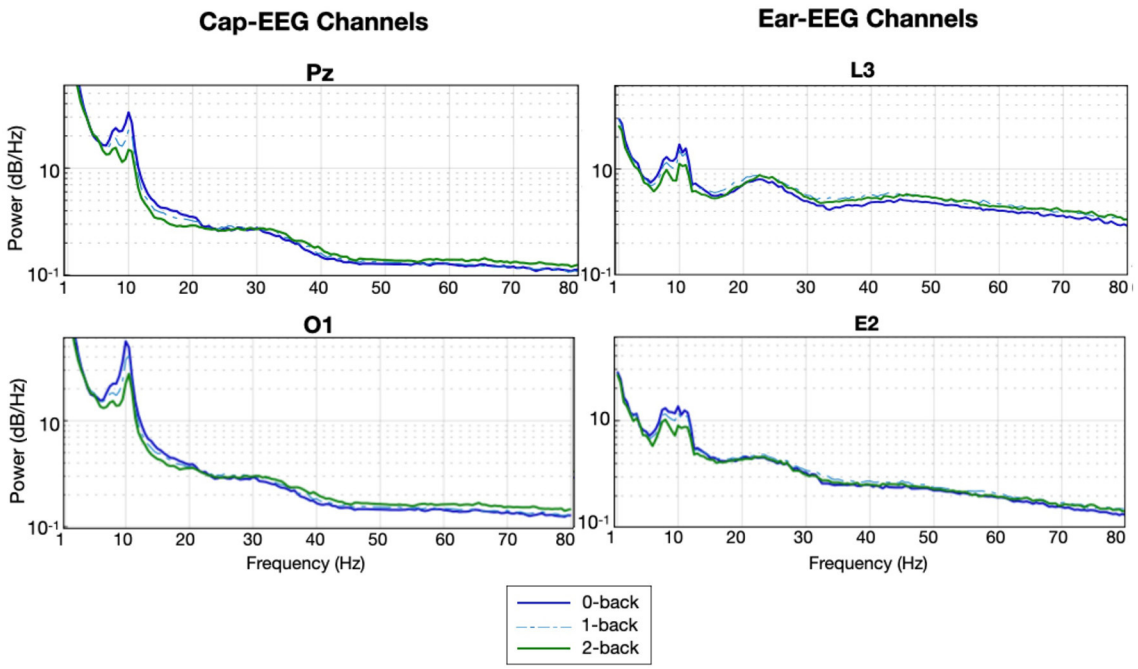


FIGURE 3
Grand average PSD for the N-back task for selected cap and ear-EEG channels (N = 14).

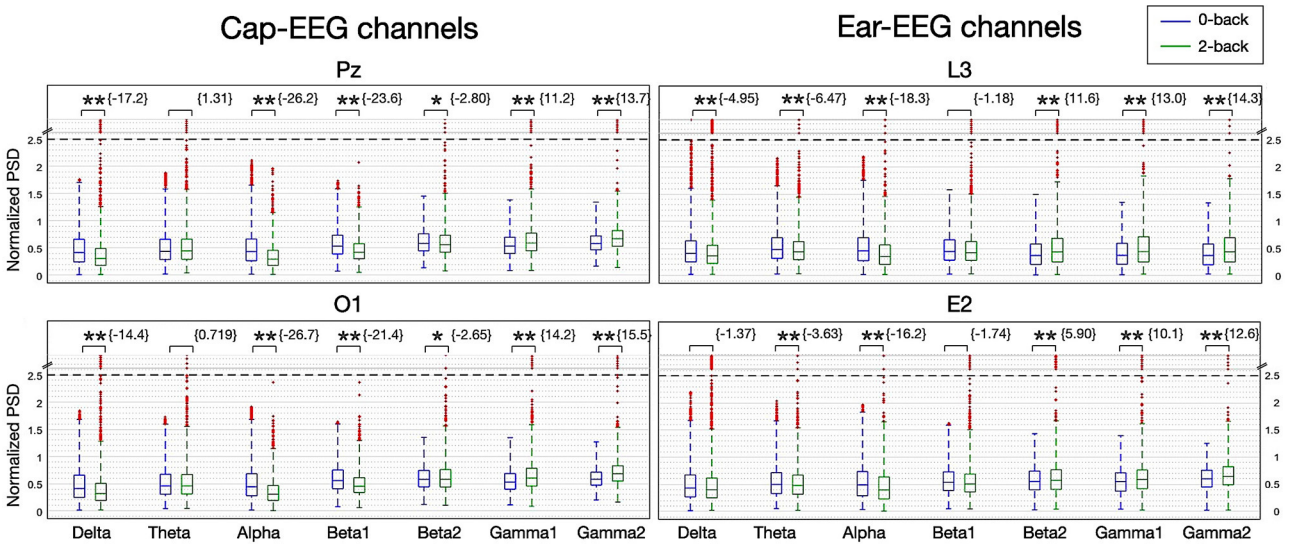


FIGURE 4
Boxplot of PSD for seven frequency bands for the N-back task for selected cap and ear-EEG channels (N = 14; *z-score>±2.3262 eq. p = 0.01, **z-score>±3.0902 eq. p = 0.001).

the cap channels and negative for the ear channels, only the most demanding condition distributions have outliers, with extreme values, for channels over both recording systems for these three highest power bands.

Spatial distribution of the PSD over the seven frequency bands can be observed in Figure 5, in which the first two rows show grand average EEG activity for the 0-back and 2-back condition. The last row represents the difference between both conditions by mapping the z-scores obtained through permutation testing for each frequency band at each electrode location. Blue regions represent a decrease in EEG activity with increasing mental load for this neural oscillation; red regions,

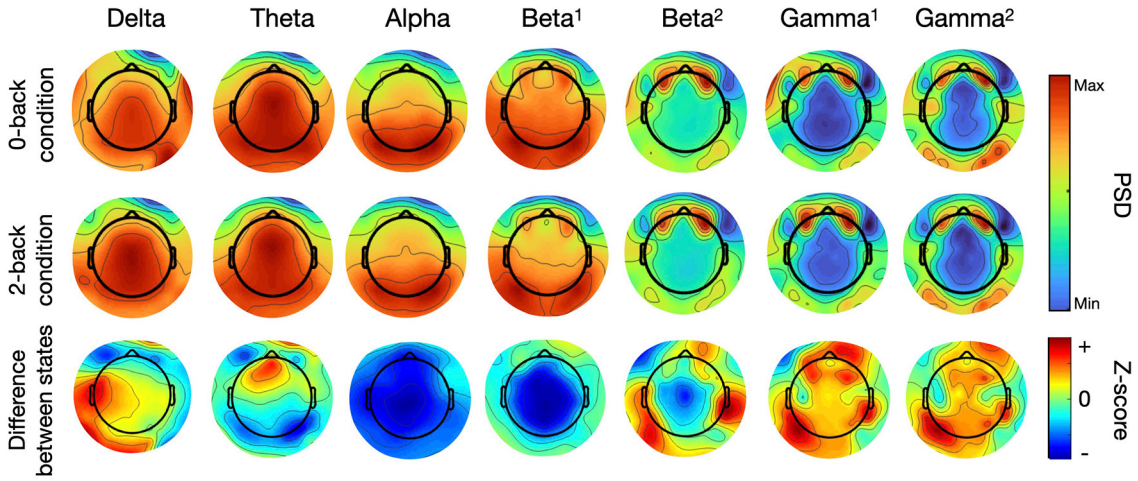


FIGURE 5
Topographical maps of PSD for seven frequency bands for the N-back task (N = 14, first two rows are grand averages, the last is the statistical results of permutation testing between both conditions–positive z-scores (red) indicate an increase in PSD from the 0-back to the 2-back, negative z-scores (blue) indicate a decrease).

an increase. The intensity of the color conveys the degree of statistical significance.

The grand average maps for each condition at frequency ranges under 20 Hz - delta to beta 1–concentrate high PSD values over the center of the map with a more posterior focus for the alpha band and the beta 1. Higher frequency range maps show that PSD values are the highest over the ocular regions, due mostly to electro-oculography (EOG) and other muscle-related artefacts. This activity was highlighted by the presence of two dedicated eye electrodes, placed under each eye. Other regions of interest for these higher frequency bands can be seen over the occipital region, increasing seemingly with each neural oscillation range from beta 2 to gamma 2.

When comparing the activity between both working memory conditions, the difference maps show a decrease in activity (represented in blue) for the lower frequency bands. The alpha and beta 1 difference maps are predominantly blue over the entire scalp, while the difference map from the theta band exhibits a decrease in activity in the occipital region that was not visible in the previous maps. Regarding the 3 higher frequency bands, beta 2 (20-30Hz), gamma 1 (30-50Hz) and gamma 2 (50-100Hz), although PSD values were higher over the ocular regions, the differences in PSD distributions across the scalp point to an increase (represented in red) in activity between conditions distributed over multiple regions spanning the entire scalp. Interestingly, regions of most significant statistical difference are concentrated at the back and to the sides of the topographical maps, indicating that the differences between working memory conditions seem to emerge from an increase in EEG activity distributed over the parietal and occipital lobe and not driven by artefacts. These results are in line with our previous findings for the different neural oscillations.

Figure 6 shows the differences at an individual subject level. The heatmaps are the result of multiple statistical analyses conducted for each subject and later summed together to highlight which PSD regions and which channels are statistically significant when comparing high and low working memory conditions. At the individual level, subjects exhibited strong differences throughout the power spectrum, which were assessed through permutation testing for each 1 Hz interval bin for individual cap and ear channels. For each subject, every bin was tested and given a binary result, 1 if the difference in PSD was significant at an equivalent p-value of 0.01, 0 if it was not significant. Converting the signed z-score to their p-values allowed positive and negative differences to be considered significant instead of canceling their effect when taken in a grand average (Figure 3) or a distribution (Figure 4) approach. Indeed some subjects had inverted effects when considering the same frequency bin and channel. Additionally and to account for the considerable amount of tests per subject (100 bins for 12 channels), p-value threshold were adjusted for each subject using the FDR method. Corrected significance threshold actually varied between 0.001 and 0.009 for the cap-EEG channels and from 0.00006 to 0.008 for the ear-EEG channels. Each rectangle of the statistical heatmap represents the sum of subjects for which this PSD bin was significantly different at this channel. Twelve channels were represented for each recording system, the same as were later used for classification. The alpha peak range, centered around 10 Hz, is visibly lighter across all channels considered, for both recording systems. Some other areas are more significant for the cap-EEG channels: the delta range (1–3 Hz) and the frequency bins right after the alpha peak where we can distinguish a slow gradient of color throughout the low beta range from 12 to 20 Hz ending with a darker, less differentiated EEG activity between 20 and 30 Hz. For the gamma range, beyond 30 Hz, both recording systems indicate that a high number of individual subjects exhibit significant differences over this frequency range.

Figure 7 shows the grand average ERP waveforms generated by the N-back task for the same electrodes as the PSD figures. Cap-EEG channels showed peaks at much higher amplitudes than ear-EEG channels. Given that the ear-EEG ERP is considerably noisier, a 4-point moving average was applied to smooth the data. For the

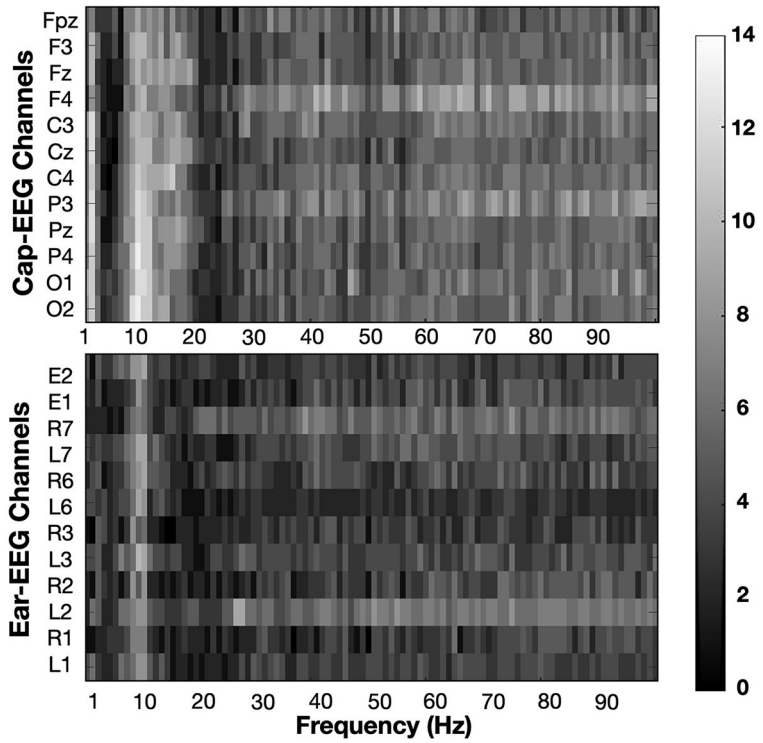


FIGURE 6
Heatmaps show the sum of subjects (from 0 to 14) for which PSD values are significantly different between working memory condition (0-back vs 2-back). Significance threshold was placed at the subject level at a FDR corrected z-score of ± 2.3262 (equivalent to $p = 0.01$).

cap-EEG channels, a distinct P300 response started around 250 ms and peaked at around 450-500 ms after stimulus onset. The P300 peak was visibly higher for lower working memory conditions and return to baseline was faster. The ear-EEG waveforms are not as clear. The around-the-ear electrode L3 highlights a visible P300 response starting a bit earlier than for the cap-EEG channels but the differences between conditions are not clear at the P300 peak level, return to baseline seems to be faster again for lower working memory condition. The in-ear data grand average ERP shows a clear P100 response, present in all the considered channels, but the P300 response is not visible. This difference between recording equipment, between the cap-EEG data's recognizable P300 peak and the ear-EEG data's much less discernible ERP waveforms is supported by the statistical heatmap for the temporal data analysis, Figure 8. This heatmap was obtained using the same method as for Figure 6, comparing at the individual subject level, PSD for low (0-back) and high working memory (2-back) conditions at each time-point (10ms intervals) and for each of the twelve channel subsets. Significance threshold at the subject level was place at an equivalent of $p = 0.01$ and corrected for multiple comparisons using the FDR method, actual p-value thresholds varied between 0.002 to 0.005 for the cap-EEG channels and from 0.000003 to 0.0014 for the ear-EEG channels. Each rectangle of the heatmaps represents the sum of subjects for which this amplitude is significantly different at this channel. The cap-EEG channels showed very bright regions around 500 ms, corresponding to the peak of the P300 response seen in the grand average waveforms. However, the ear-EEG channels, particularly the in-ear channels (E1 and E2) don't show as much shared differences in the temporal domain. Some outer-ear channels were still able to detect significant changes in amplitude in the range of the P300 response for a majority of subjects.

3.1.2. Arithmetic task

For this task, the PSD was taken from 2-second epochs taken at regular interval, no time domain analysis was performed. To assess the fluctuations in EEG activity between a high and a low cognitive workload, subjects were asked to solve arithmetic problems with varying degrees of difficulty. The differences were assessed on group data from ten subjects and an individual subject basis to create the heatmaps.

Figure 9 shows the grand average PSD for the arithmetic task. For this task also the different mental load conditions can be differentiated on these plots for both recording systems. The alpha peak was harder to differentiate between low and high workload conditions for the cap-EEG than for the ear-EEG. For this task and for both recording equipment, the more demanding condition appear to generate higher PSD than the less demanding one, over the entire frequency range considered. The grand average ear-EEG PSD difference is quite striking from the alpha range forward. The gamma band PSD for the cap-EEG data was also distinctly higher for the more demanding workload condition.

The boxplots of Figure 10 represent the distribution of PSD for all subjects for the seven frequency bands. The data was analyzed in the same manner as for the N-back task, the frequency bands were the same, the normalization procedure and statistical

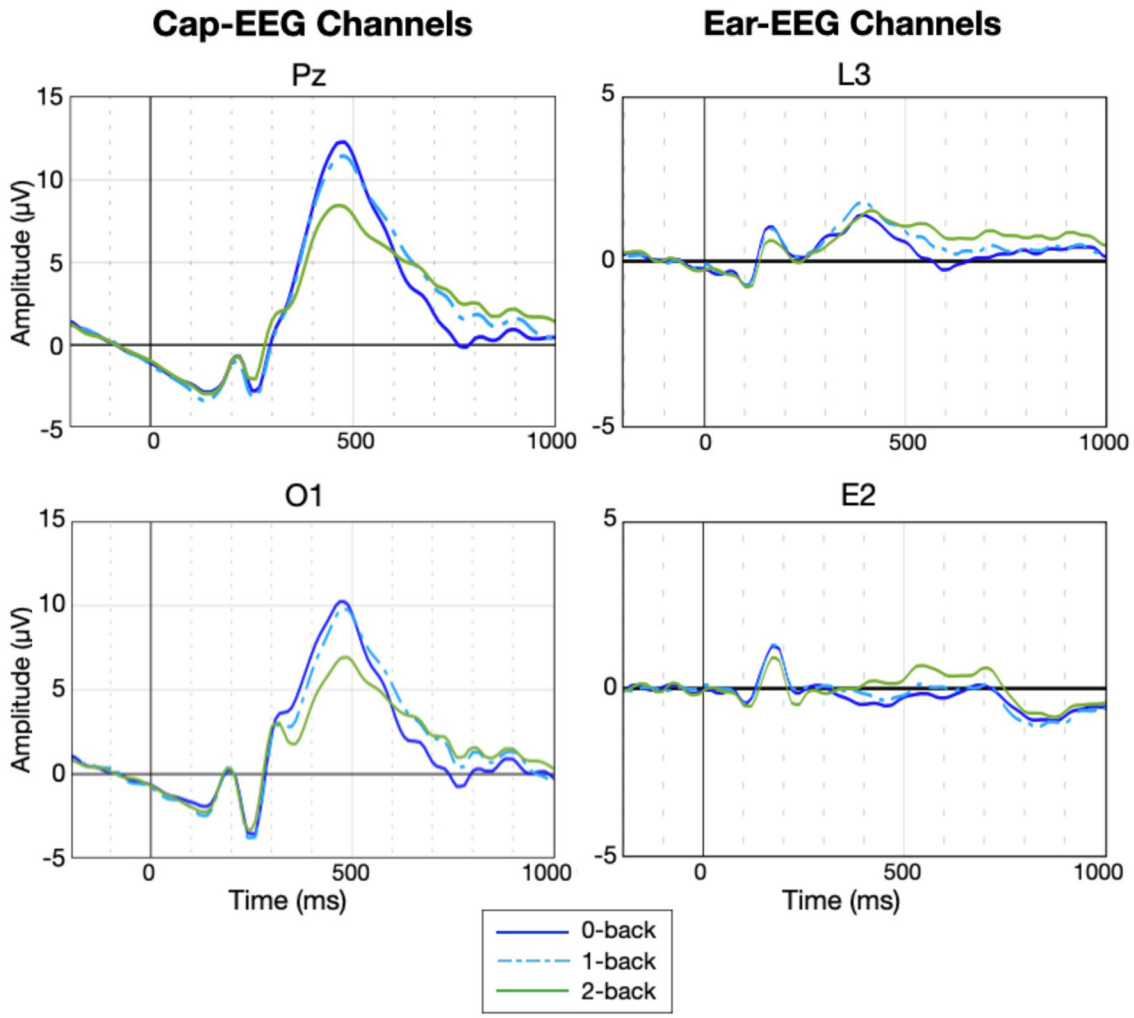


FIGURE 7
Grand average ERP for the N-back task for selected cap and ear-EEG channels (*N* = 14).

approach were also the same (refer to Figure 6). The boxplots confirm and highlight that the more demanding cognitive workload condition's PSD is statistically higher than the less demanding condition for all frequency bands and for all channels considered from both recording equipment, indeed all z-scores obtained from permutation testing are highly significant and they are all positive. Despite the normalization of the data and the increase in the percentage of distribution included within the whiskers of the plots, there is still a large number of outliers for this task, like for the one before. For this task however, extreme value outliers are only found for the more demanding condition, and the higher the frequency bands, the more this difference in outlier presence differs between the conditions. From beta 2 to the gamma ranges, the distributions for the low workload condition show no outliers for all channels considered over both recording systems while the high workload condition distributions exhibit many and always with some extreme values. These non-uniform distributions could suggest that the pooled data from all subjects might need better processing methods to harmonize results and lower variance.

The topographical mapping of the seven frequency bands is shown in Figure 11. The maps show the grand average PSD at each channel location for both conditions and the difference between conditions by way of projecting the results of a statistical analysis at each electrode location (cf to Figure 5 for more details as the procedure is the same). The grand average maps are not so informative, there seems to be a strong unilateral posterior component present in all frequency ranges and strong activation over the ocular region as well, due to EOG artefacts, for the beta and gamma bands. Other important regions and differences between states are not easily identifiable in these rows but they come into focus with the difference maps. Mostly the difference maps show an increase in EEG activity (seen in red) throughout the frequency range considered. The posterior region, over the occipital lobe, seems to be the most significant when considering the highest (high beta and both gammas) and the lowest (delta and theta) frequency bands, indicating again that differences between conditions is emerging from actual EEG activity. Maps from the middle of the frequency range, alpha and low beta, show more lateralized regions of interest over the parietal and frontal lobe.

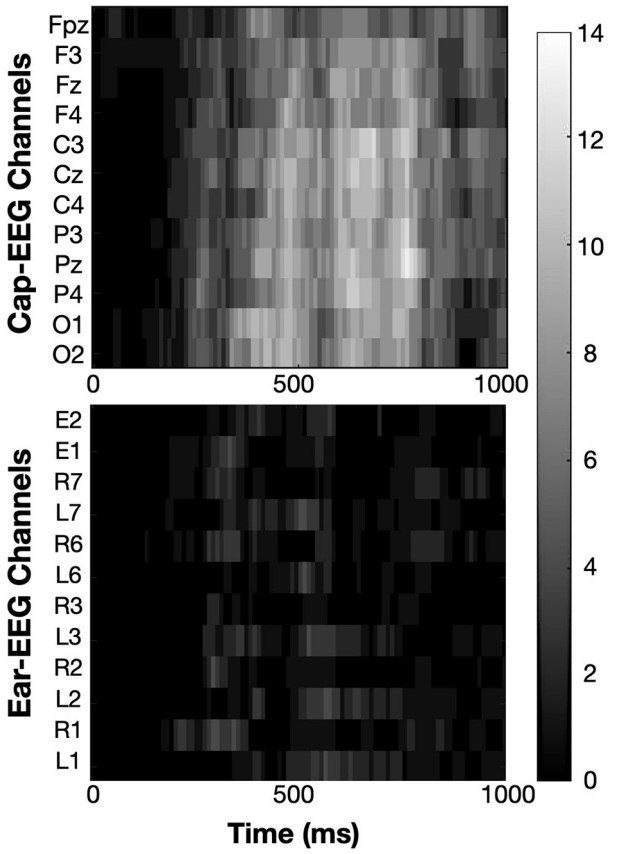


FIGURE 8
Heatmaps show the sum of subjects (from 0 to 14) for which ERP amplitudes are significantly different between working memory condition (0-back vs 2-back). Significance threshold was placed at the subject level at a FDR corrected z-score of ±2.3262 (equivalent to p = 0.01).

For this task, EOG activity seems to have a localized differentiated effect between mental load conditions, particularly for the higher frequency bands where artifact activity decreases.

Figure 12 highlights the number of subjects showing a significant difference between low and high cognitive workload conditions in PSD from 1 to 100 Hz for the same twelve channels subset for each recording equipment as before. The approach is the same as for the N-back task, refer to Figure 6 for more details. Significance was calculated through permutation testing and the resulting statistical scores adjusted using the FDR method for multiple comparison correction. Significance threshold was put at p = 0.01, adjusted p-values varied for each subject ranging from 0.00011 to 0.0072 with an average of 0.0031 for cap-EEG channels and from 0.00017 to 0.0085 with an average of 0.0037 for ear-EEG channels. The alpha range for this task appear less differentiated across conditions than for the last. Some channels have some very bright bins in this interval, but for the cap-EEG data at least, there is no clear band across all channels. The ear-EEG data has a clearer indication of strong differences in the alpha range flanked by a very dark region over the theta range (4–7Hz) and a darker region just following it. This is not to say that there were less significant differences in the PSD for this task, very light areas are found throughout the heatmaps. They were more largely distributed across the higher frequency ranges, mostly beyond 20 Hz for the ear-EEG data and beyond 30 Hz for the cap-EEG data. Lastly, the cap-EEG data appears to be less significant for individual subjects beyond 80 Hz although this tendency was not observed in the ear-EEG data.

3.2. Classification results

3.2.1. N-back task

Figure 13 shows individual subject's single-trial classification accuracies obtained using an swLDA classifier to predict the class of individual trials between the high (2-back) and low (0-back) working memory conditions. The algorithm first performed a feature selection step, the step-wise linear regression which kept only the more statistically differentiated features. Then the model was trained on a portion of the trials and tested on the remaining trials using only the selected features. Accuracy results show the percentage of correctly classified trials within the testing sets using a 10-fold cross-validation scheme. For each subject, the average of these 10 accuracies was reported and the average over all subjects is plotted in the same figure. The features used were extracted from the 1 Hz spectral feature range model. This model consists of PSD extracted from each epoch or trial from 1 to 100 Hz in 1 Hz interval bins, much like the statistical heatmap data (Figure 6). The heatmap figure represented the same channels considered for classification, the lighter regions of the heatmaps showed where the significant features are concentrated and there is a strong overlap between these regions and the selected features of the classifier. They are the features that help differentiate between working memory loads. This feature model started with 100 features (PSDs) per channel, 1,200 for the 12 channel model, which represents a higher number than the number of trials (768). This creates issues of dimensionality that are problematic for classification purposes. The ill-advised feature to sample ratio was mitigated through the feature selection regression. On average, over fourteen subjects, the final number of features kept for this model went from 1,200 for the twelve channel models to 306 for the cap-EEG data and 285 for the ear-EEG data. For the two channel models, the initial number of features was 200 and the number of features used for classification was on average 35 for the cap-EEG data and 34 for the ear-EEG data.

Classification results of Figure 13 show that all subjects achieved significant above-chance level performance for both recording equipment, for the twelve and the two-channel models. For the twelve channel models, average accuracy was 96% for the cap-EEG and 95% for the ear-EEG;. For the two-channel models, average accuracy was 76% for the cap-EEG and 74% for the in-ear EEG. It might appear that cap-EEG data yielded higher classification accuracies than ear-EEG data but these differences were not found to be statistically significant with a paired t-test comparing the 10 fold classification accuracies of all subjects and a p = 0.001 threshold (p = 0.034 for the twelve channels and p = 0.230 for the two channels). On an individual subject level, it is interesting to note that some of the two channel ear-EEG models which represent only dry in-ear electrodes outperform greatly the occipital channels of the cap-EEG models, it is the case of subject 1 and 6 for example and subject 4 to a lesser extent.

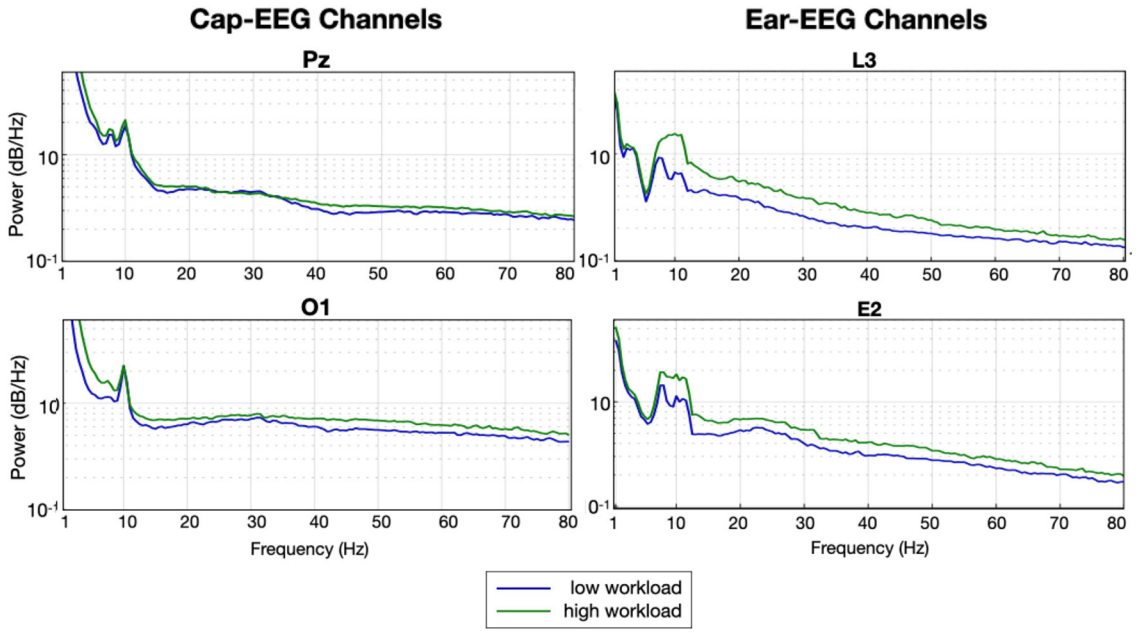


FIGURE 9
Grand average PSD for the arithmetic task for selected cap and ear-EEG channels (N = 10).

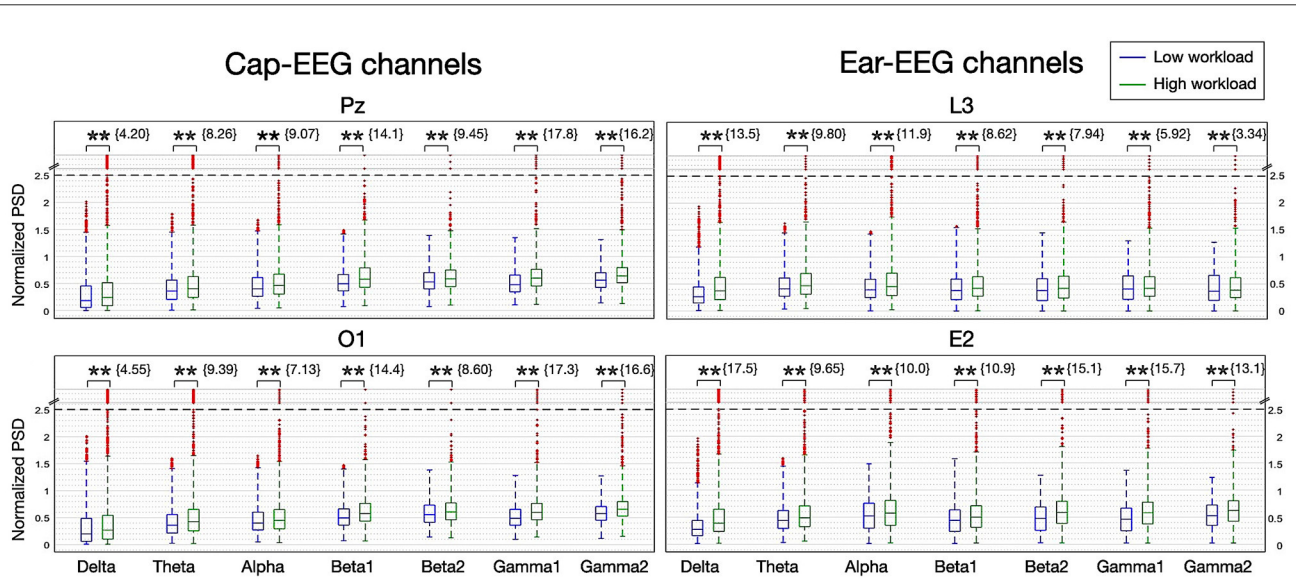


FIGURE 10
Boxplot of PSD for seven frequency bands for the arithmetic task for selected cap and ear-EEG channels (N=10; *z-score > ±2.3262 - eq. p = 0.01, **z-score > ±3.0902 - eq. p = 0.001).

For the N-back task, time domain features were also available. Table 1 provides the classification accuracy for different feature modalities: spectral, temporal and mixed. Both cap and ear-EEG data were considered for both channel models. The spectral 1 Hz interval bins feature and the temporal 1,000 ms (in 10 ms interval) features started out with the same number of initial features, 100. The mixed or "fusion" model had 200 initial features per channels. The results show that the spectral domain features outperformed the temporal domain features and that the fusion model outperformed both of them. This is statistically significant for both channel subsets across both recording systems with p-values all under 10^{-7}. While the performances for the spectral range models did not differ statistically between recording equipment. The performances of the classifier for the temporal and fusion models were all significantly better for the cap-EEG data than the ear-EEG data with p-values all under 10^{-4} for both channel subsets and both feature model.

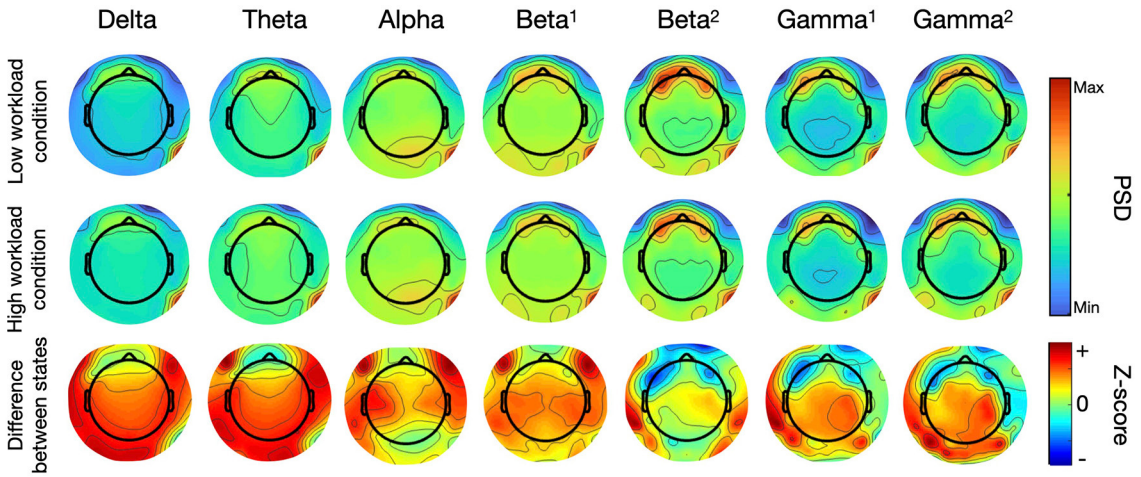


FIGURE 11
Topographical maps of PSD for seven frequency bands for the arithmetic task (N=10, first two rows are grand averages, the last is the statistical results of permutation testing between both conditions–positive z-scores (red) indicate an increase in PSD from the 0-back to the 2-back, negative z-scores (blue) indicate a decrease).

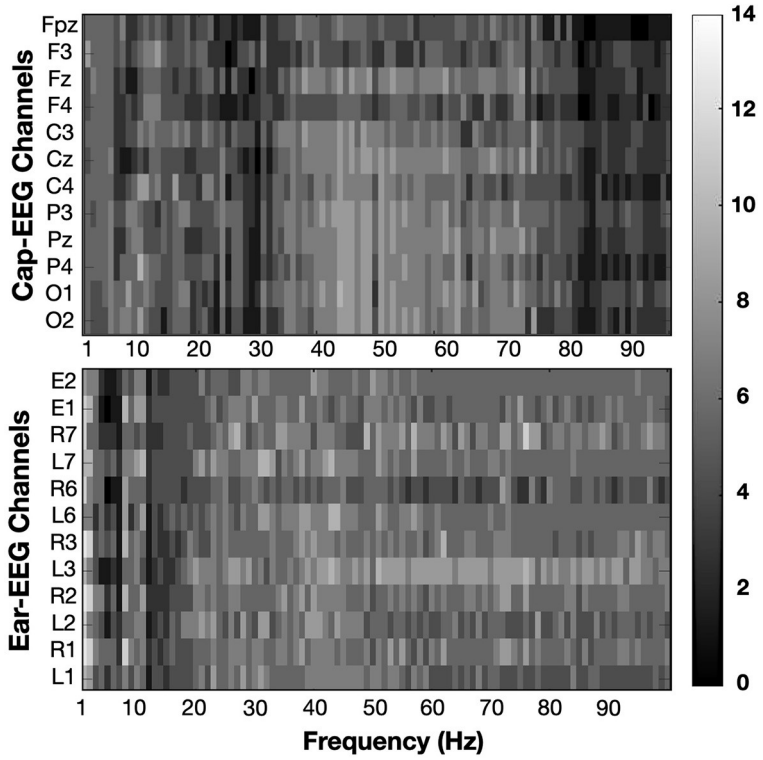


FIGURE 12
Heatmaps showing the sum of subjects (from 0 to 10) for which PSD values are significantly different between cognitive workload conditions (low vs. high). Significance threshold was placed at the subject level at a FDR corrected z-score of ± 2.3262 (equivalent to $p = 0.01$).

When considering the final number of features for each model, it can be noted that the final number of temporal features was a lot less than the final number of spectral features, about 80 % less for the 12 channel models and close to 50 % for the 2 channel model. The final number of mixed features selected by the step-wise linear regression was unexpectedly higher than the addition of those different modality features for the twelve channel configuration, 37% more features on average. This could explain the abnormally high classification accuracies of the fusion model for the twelve channel models, probably brought on by over-fitting. This was not the case for the 2 channel model.

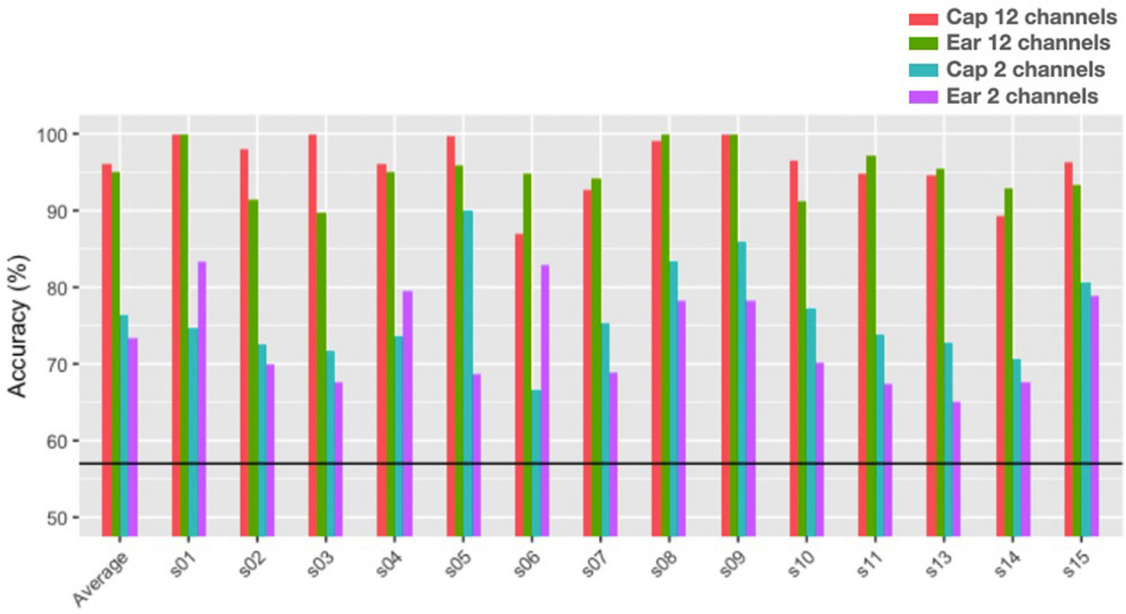


FIGURE 13
Single-Trial classification accuracies for the N-back task with a 1 Hz interval bins and a 1–100 Hz frequency range (the black line represents chance level at 57%).

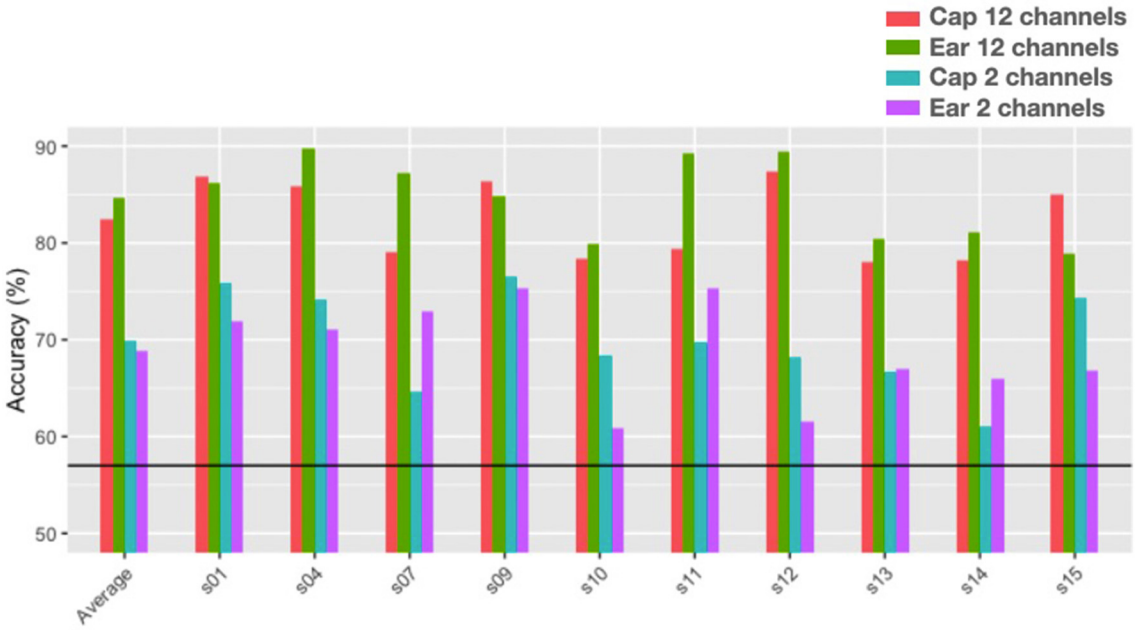


FIGURE 14
Single-trial classification accuracies for the arithmetic task with a 1 Hz interval bins and a 1–100 Hz frequency range (the black line represents chance level at 57%).

3.2.2. Arithmetic task

Figure 14 shows the single subject classification accuracy for the same model as in Figure 13 for the arithmetic task (1 Hz interval frequency range model from 1 to 100 Hz) for each recording system and for a twelve and two-channel feature set. The initial number of features was again undesirably close to the number of trials, around 1,392 compared to 1,200 initial features for the twelve-channel model. Again, the step-wise linear regression step helped reduce the number of features considered for classification to 139 for the twelve-channel cap-EEG, 171 for the twelve-channel cap-EEG

TABLE 1 N-back task single-trial classification results for different feature modalities (in grey is the final number of features kept by each classifier over the initial number of features available).

	12 channels		2 channels	
	cap-EEG	ear-EEG	cap-EEG	ear-EEG
Range 1 to 100 Hz 1 Hz steps	96 % (306/1,200)	95 % (285/1200)	76 (35/200 %)	73 % (34/200)
Temporal 1 to 1000 ms	86 % (76/1,200)	76 % (57/1200)	72 % (17/200)	64 % (16/200)
Fusion: temporal and 1 to 100 Hz	100 % (564/2,400)	99 % (592/2400)	81 % (48/400)	77 % (52/400)

data, 34 for the two-channel cap-EEG and 36 for the two-channel ear-EEG. The figure shows that all subjects performed better than chance-level for the twelve-channel feature set with an average of 81% for cap-EEG, 84% for ear-EEG. For the two-channel feature set, the average was 70% for the cap-EEG and 69% for the in-ear EEG. On average, it seems that the ear-EEG features outperformed the cap-EEG ones for the twelve channel model but not for the two channel model. However, none of these differences are significant at an alpha of 0.001 although the twelve channel ear-EEG model comes close to this threshold (p= 0.0013 for the twelve-channel, p= 0.309 for the two-channel). Some individual subjects show better performances for the in-ear EEG data compared to the cap in this task as well, such is the case for subject 7 and 11 for example.

3.2.3. Spectral feature model comparison

Figure 15 is useful to assess the influence of certain feature parameters over classification accuracies. The same tendencies can be observed for both tasks: on the left, a decrease in classification accuracy with decreasing spectral resolution; on the right, a decrease in classification accuracy with decreasing frequency cut-off for the spectral range considered. The feature models in the left plot all have a frequency cut-off of 100 Hz. The accuracy results from the right plots are those of the twelve channel ear-EEG model, the tendencies were the same for other channel subsets. The feature sets are described in detail in Section 2.5. They are labeled as 0.5 Hz for 0.5 Hz-interval bins (200 initial features per channel); 1 Hz for the 1 Hz-interval bins (100 initial features per channel); the 5 Hz model corresponds to bands of 5 Hz or less (21 initial features per channel); and finally the band models have the same number of initial features per channel as their band number.

Statistically for the N-back task, each decrease in resolution from 0.5 Hz down to 5 Hz did yield a significantly lower reported classification accuracy average for all channel models. After that resolution, drops in accuracy are not significant (alpha 0.001) from one model to the next until reaching the 5 bands model where all four channel models perform statistically worse than all other models considered. The drop in accuracy and subsequent stagnation can be linked to the number of features in the final models. For the twelve channel models the final number of features is about 540 for the 0.5 Hz models and just under 300 for the 1 Hz model. After this, it drops to below 50 features for the 5 Hz and 17 for the 5 band models. The two channel models only have 5 or 6 final features considered by the classifier yet they still manage to yield accuracies over chance-level (73% for the cap-EEG and 65% for the ear-EEG).

For the arithmetic task, statistical analysis results are quite similar. Significant drops in accuracy can be found from 0.5 to 1 Hz and then again from 1Hz to 5Hz for all four channel. After this, no significant drop from one model to the next is recorded until the 5 band model which performs significantly worse than the 5 Hz model but only for the twelve channel models.

The right side plots show that classification accuracy decreases when the gamma cut-off frequency is lowered. The seven band model stops at 80 Hz because with a 45 Hz cut-off frequency, it consisted of the same features as the 5 Hz model. All reported average accuracy results plotted for both tasks and for the 5 feature models were lower when the cut-off frequency dropped from 100 to 80 Hz and from 80 to 45 Hz. However, not all decreases were statistically significant. Lowering the cut-off to 45 Hz was significant for both tasks for the three range models considered (0.5, 1, and 5 Hz); while the drop in accuracy going to a 80 Hz cut-off frequency was statistically significant for the 0.5Hz and the 1Hz model when considering the N-back task; and for the 1Hz model for the arithmetic task.

4. Discussion

Future BCIs have the potential to change how we interact with technology, assuming they can be used outside of controlled environments. Besides comfort and aesthetics, the signal must be of sufficient quality if one is to reliably decode different mental states. Mobile EEG recording equipment with proven signal quality is already available (Debener et al., 2012; De Vos et al., 2014a). In this study, we investigated the potential of ear-EEG to monitor focus, a concept that overlaps with notions such as concentration, attention, engagement and intensity of a task. Tasks were chosen to represent different features of focus, namely working memory and cognitive workload. The goal in both of these tasks was to differentiate between low and high mental demand to ascertain differences in EEG activity that could be indicative the level of focus. We compared the quality of neural oscillation recordings of a mobile ear-EEG recording system to that of a conventional 84-channel wet cap-EEG wired system, used concurrently. The ear-EEG electrodes were wet around-the-ear electrodes (cEEGrids) and dry in-ear electrodes (TIPtrodes^{TM}).

When comparing the cap-EEG data obtained with existing studies using similar paradigms (Rebsamen et al., 2011; Brouwer et al., 2012), our results were comparable. For the N-back task, we found a strong and significant alpha peak with its power inversely proportional to the working memory load (Figure 3). The peak was visible in both cap and ear channels, although cap channel power was stronger for this interval. In the higher frequency ranges, above 30 Hz, it seemed the PSD order for both recording equipment inverted with the higher demand translating into higher PSD levels over high beta and gamma bands. Over this higher frequency range, ear channels recorded slightly stronger PSD signals. Overall, these grand average power spectrum seem to offer information of comparable quality across both recording systems as further evidenced by Figure 4. This next figure highlights the differences in distribution of all frequency ranges in more compelling detail,

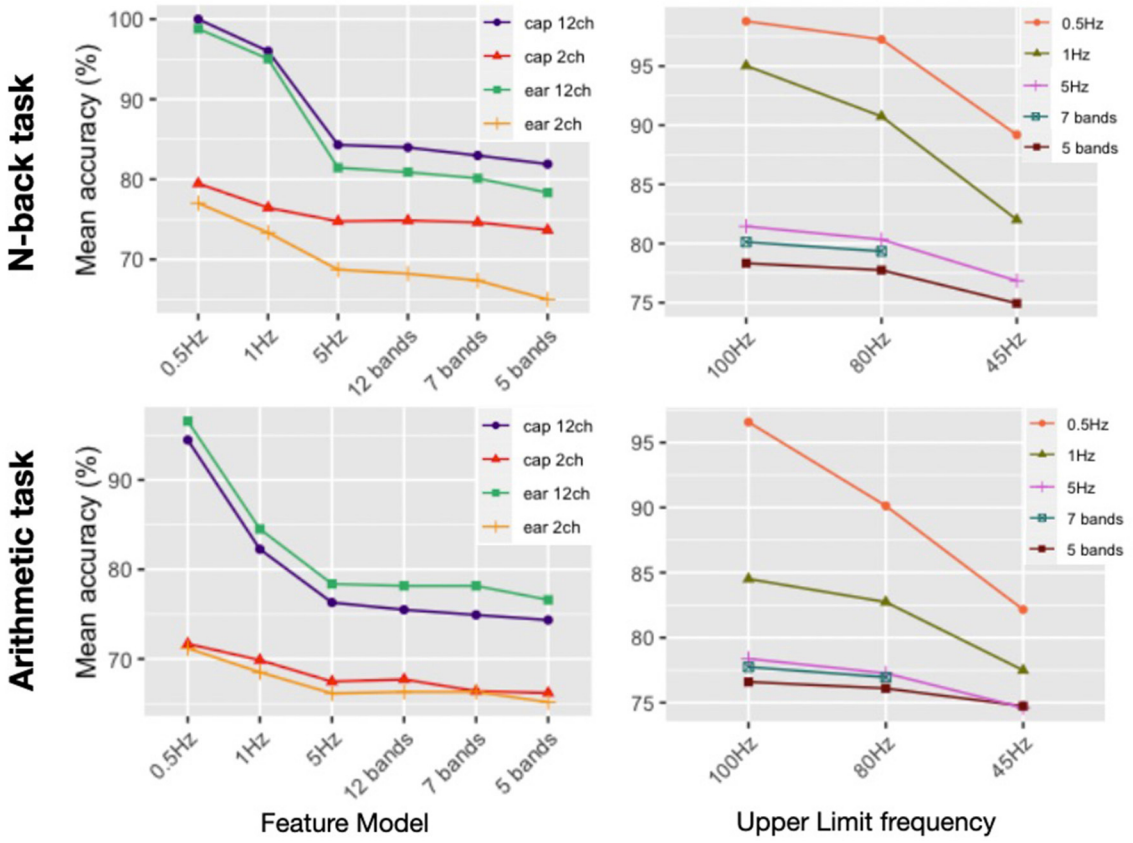


FIGURE 15
Average of individual subject's classification accuracies for different spectral resolution and frequency range cut-off for the N-back (N = 14) and arithmetic task (N = 10).

through its representation of the PSD distributions but also thanks to the statistical results it features. The results of the boxplots establish that for this task, EEG activity elicited from a higher working memory demand translates into statistically lower PSD for frequency ranges under 20 Hz; from delta to alpha. Above 30 Hz, higher working memory demand translates to a higher PSD. The transition between this inversion occurs over the beta range from 12 to 30 Hz where it does fluctuate between channels and range considered. Cap channels are quite effective at showing differences in the lower beta range, behaving much like the alpha range distribution: lower medians for the more demanding condition and less outlier values. For the higher beta range, the around-the-ear channel show a statistically more significant and visually more compelling difference where the more demanding condition show a higher distribution and more outliers. The dry in-ear EEG channels' results seemed less conclusive when considering the grand averages, particularly when it came to the higher frequency bands, but the differences evidenced in the boxplots were still very significant for the gamma frequency range.

The topographical maps show how these differences in PSD were distributed spatially. They are consistent with the previous results, upholding the choice of these selected channels to represent EEG activity for this recording system. The shift between the lower frequency ranges which were predominantly blue, indicative of a decrease in activity to maps with a lot more red, indicative of an increase in EEG activity, for the higher frequency range is complemented with more information on the spatial origin of these differences in PSD. The higher frequency bands (high beta and gamma oscillations), residual artifacts of eye movement clearly contaminated the PSD topographical maps. However, it was shown that activity in this ocular region is not where the differences between working memory load are the strongest. For these higher frequency bands, higher working memory load had the strongest statistical significance over posterior regions of the maps. Visible clusters of heightened activity can be seen over the temporal lobes reaching more towards the occipital region. This lets us surmise that although PSD in the high beta and gamma ranges incorporates residual muscle activity which contributed unwittingly to the results, the differences in mental states are in fact driven by activity emanating from cortical areas. This was important to show particularly for the gamma range where EOG activity is known to persist even after appropriate data pre-processing and can be underestimated when using a nose reference (Yuval-Greenberg et al., 2008).

Heatmaps of Figure 6 have shown that differences at the individual subject level are strong for the alpha range and for the higher frequency bands (30Hz and up), sometimes stronger

than grand average or cumulative distribution differences. Boxplot representation showed that there is a high degree of variance in data sets obtained from multiple subjects and that the medians are not the best tool to assess differences between these populations. Heatmaps considered individual differences, which were sometimes of opposite signs, to assess which frequency ranges were the most indicative of change in EEG activity between mental load conditions.

The classification accuracy results for this task obtained with a spectral range feature set from 1 to 100 Hz in 1 Hz intervals were well above chance-level, that is, over 95% for the twelve-channel models and over 70% for the two-channel models. The working memory condition could reliably be decoded from both cap and ear-EEG signals as well as in-ear EEG. Cap and ear EEG recording equipment produced comparable results and although the cap-EEG feature sets often had a few percentage points over ear-EEG feature sets, this difference was not significant. This was all the more impressive given that the twelve-channel cap-EEG electrodes were placed around the entire head versus the ear-EEG electrodes closer together and condensed in two much narrower regions around and inside the ears. The two-channel model's comparable performances were noteworthy particularly since we compared occipital wet electrodes to in-ear dry electrodes for a task with visual stimuli. On an individual subject basis, none of the recording equipment was consistently better than the other. A better result with the twelve-channel cap-EEG feature model did not mean that the two-channel cap-EEG would perform better and vice versa.

For the arithmetic task, again we observed clear differences between mental load conditions, however these were not always the same differences as what we saw for the N-back task. The alpha range provided a visible peak and could be used to distinguish between conditions but the difference was mostly for the ear-EEG channels this time and said difference was in the opposite direction than for the other task. The cap-EEG channels grand average representation of PSD (Figure 3) did not provide a clear differentiation between workload levels at this frequency range, although significant differences were confirmed with the boxplots (Figure 4). This is in line with the concept of internalized attention vs. externalized attention (Magosso et al., 2019). For this task, it was observed that with a higher workload came a higher PSD for the entire frequency range considered for both recording equipment. Only the high beta range (20 to 30 Hz) showed areas of overlap between conditions and for cap channels only. Other than this, the low workload conditions had systematically and significantly lower PSD at every frequency range considered. For a majority of the bands, the standard deviation for the low workload condition was also much smaller than the high workload condition. Although the differences in EEG activity for the lower frequency bands (under 20 Hz) were not what was observed for the N-back task, the gamma band distributions had important similarities between the two tasks. When considering the topographical distribution of PSD for each neural oscillation (Figure 5), the difference maps all showed large areas with positive z-scores, which translates to stronger PSD with higher workload demand. The regions where these differences were the strongest and most significant were distributed over several cortical regions according to the frequency range considered. The alpha and lower beta band's PSD changes were concentrated over the temporal region which could be an indication as to why the ear-EEG channels performed better at differentiating activity in that range. A phenomenon confirmed again in the statistical heatmap of Figure 12, where the alpha band was visible for the ear but not so much for the cap-EEG channels. Regarding the higher frequency range maps, similar conclusions could be drawn for both tasks for high beta and both gamma bands. Specifically, grand average workload condition maps pointed to the ocular region, indicative of muscle artefacts, but the most significant areas when considering differences between conditions pointed to the parietal and occipital regions consistent with actual EEG activity. Some clusters were also found closer to the temporal regions for the beta bands. These considerations are important since the features used for classification are selected based on their statistical difference between conditions. Conclusions of the topographical analyses therefore support the argument that the features used for classification are in fact related to actual brain activity dynamics. The classification accuracy results for this task, obtained with the same range model as before, were again well above chance-level, 80% and more for the twelve channel models and close to 70% for the two channel models. The cognitive workload could be decoded from both cap and ear-EEG signals, around and inside the ear. Statistical analysis of both task, did reveal that for this task, the ear-EEG channel sets performed slightly better than the cap-EEG channels for some subjects over both channel models, twelve and two.

We conclude that EEG spectral features recorded from inside and around the ear can reliably assess different focus levels may it be related to working memory or cognitive workload features. These features systematically allow for significantly above-chance level classification accuracy between all levels of difficulty. Therefore, it would be feasible to develop a concentration or focus-monitoring BCI in the form of an earpiece.

The next steps for this line of research will be to test these conclusions under less restrictive conditions, outside of a controlled environment and with more diverse and complex tasks. Such data obtained from more realistic focus conditions will obviously be noisier, and will be harder to appropriately label for the purpose of training new classification algorithms. On the other hand, the tasks presented for this study might have been biased or misrepresented the focused state of individual subjects. The degree to which the participant's might have felt involved in these repetitive tasks throughout the recording and the perceived difficulty of the task could vary as several subjects did comment on the tediousness of the N-back task in particular. Recording participants in more realistic focus scenarios might actually provide better, more consistent differences across PSD features between tasks and between participants. The nature of the tasks will most likely have an impact on which spectral ranges are best to characterize focus. As we have seen here, the alpha frequency although very effective to distinguish between working memory conditions and between cognitive workload conditions behaves in a completely opposite fashion from one task to the other. It might be too task dependent to be useful in this context. Higher frequencies in the beta and gamma bands, although mixed with muscular activity, might procure more consistent markers of focus across a variety of tasks.

This study showed grand average EEG results and distributions from all participants but reported classification accuracies for within-participant trial data. This choice was supported by Aroudi et al. (2016)'s results on attention decoding, which showed that given sufficient individual data, individual models will outperform the group model given that inter-subject variance is always larger than variance within individuals for this type of data. Nevertheless, according to the type of application considered it might still make sense to try and improve group model classification. Going back to the boxplot from both tasks, distributions of PSD values were non-uniform, skewed and contained a lot of outliers. The medians were often very close in value while the spread and reach of the distributions were in fact very different. This might be related to the high variance across subjects which could be improved through better standardization methods when comparing PSD values. Spectral selection, filtering and the use of band ratios seem like interesting avenues to address this (Berka et al., 2007). This will most likely imply further research on feature extraction procedures and preprocessing specifications that can have a great impact on classification accuracies (Farquhar and Hill, 2013). The choices made, that led to the specific features used for this study, included montage selection (electrode discrimination from eighty-four to twelve for the cap-EEG data and twenty to twelve for the ear-EEG data), resampling and filtering, epoch duration selection and base-lining. ICA was used to remove mainly the eye blink components from the data before choosing the best re-referencing option for both recording systems. The ear-EEG data provided the best results using a contralateral reference for each ear channel, this increased the distance between the reference and the channels. For the cap-EEG data, the nose reference was kept as it was the furthest from strong noise contamination from the concurrent recording systems which affected several cap-EEG channels around the ears at different locations for different participants and in intermittent fashion. Linked mastoid reference was rejected because of this unpredictable noise, most likely originating from the ear-EEG equipment. It might also have had an impact by propagation on the use of a common average reference which strongly impacted individual subject differences and PSD distributions at individual channel locations, making it harder to keep a consistent channel selection over all participants and tasks. The nose-reference although not well suited for source localization (Mahjoory et al., 2017) produced more consistent results for inter-subject comparisons of specific electrode locations. Topographical representation of PSD can be biased by a nose reference because of the proximity with ocular artefact location (Trujillo et al., 2005), therefore conclusions regarding topographical data from this study were tested against common average reference topographical representations and held up.

Stronger classification algorithms might be necessary to identify the important features shared between individuals and normalization of the spectral data will most likely be a determining factor in improving this process. In order for the algorithms to be able to generalize across individuals, selecting which frequency ranges are used as features could be very relevant as well as better standardization of the features across subjects through the use of frequency band ratios for example. On model performance comparisons, we have learned from this study that higher spectral resolution tends to produce better classification results at the individual subject level, as does a higher cut-off frequency. However, increasing spatial resolution increases the dimensionality of each sample increasing the risk of over-fitting. This explains why even though the 0.5 Hz resolution models performed better than the 1Hz resolution, the later models were used for visualization and for single trial classification since the final number of features was more appropriate considering our sample size. The choice of a 100 Hz cut-off frequency was influenced by the feature resolution as this model performed significantly better with this cut-off frequency for both tasks. It might be interesting to do future research on maximum useful frequency cut-off as ear-EEG channel features in particular seem like they would be significant past this 100Hz limit. Lastly, the study of temporal features showed that even though they were capable of predicting with above chance-level accuracy the levels of cognitive demand for both recording equipment and channel subsets, they performed worse than spectral ones. Results of this study have shown that while ear-EEG spectral features performed on par with cap-EEG spectral features, ear-EEG temporal data is much noisier than cap-EEG's temporal data as evidenced by the grand average ERP waveforms (Figure 7) and the temporal heatmaps (Figure 8). Dry in-ear EEG channels did not produce a grand average P300 response but were sometimes better than selected cap-EEG channels at differentiating PSD between conditions. Hence ear-EEG might be better suited for applications that rely on spectral data rather than temporal responses, such as focus monitoring. It is worth noting that mixing temporal with spectral features did improve classification results compared to spectral features alone. Although given the abnormally high performance of this type of fusion model (near 100 %), and the high dimensionality of these models, over-fitting is suspected to have happened, despite feature reduction. This will be decisive if these results are to be exploited in a BCI context, it would also make sense to report information transfer rates along with classification accuracies.

Future research should also focus on improving ear-EEG electrode technology. The TIPtrodes^{TM} are very sensitive to movement, which makes them single-use and susceptible to various sources of electromagnetic noise. Around-the-ear electrodes used in this study were wet, innovation in electrode material and treatment could allow their use as dry electrodes. Movement-tolerance of such ear-EEG recording system will surely be a research area of interest going forward. Tighter contact, possibly achieved through custom-fitted earpieces and shielding technologies, could effectively decrease the potential impact of movement artifacts when developing these applications for daily life situations.

5. Conclusion

Focus can be assessed using EEG spectral features that were recorded from sites in and around-the-ear using a mobile EEG amplifier. These features are consistent with known characteristics of working memory and cognitive workload. Ear-EEG is a promising candidate for future BCI and brain-monitoring applications, which could help

with our day-to-day activities, such as workplace security and productivity.

Data availability statement

The dataset was meant for use in this study and not for further studies. Requests to access the datasets should be directed to gcretot@critias.ca.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of Oldenburg 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

SD, MD, and JV devised the project and the main conceptual ideas. SD validated the experimental paradigms proposed by MB and GC-R. GC-R carried out the experiment and data processing, performed the analytic calculations, developed the statistical analysis and the classification algorithms, and drafted the manuscript and the figures. MD outlined the theoretical formalism and analytic calculations. MD, SD, and MB verified the article's content and added to the analysis and interpretation of the data and conclusions and together with JV they reviewed the final manuscript. JV and MD were co-supervising the project. All authors contributed to the article and approved the submitted version.

Funding

The authors declare that this study received indirect funding from EERS Global Technologies Inc. through the National Science and Engineering Research Council (NSERC) Alliance Grants program (ALLRP 566678-21). The funders were not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Acknowledgments

GC-R and JV would like to acknowledge the financial support received from NSERC Alliance (ALLRP 566678-2021), MITACS IT26677 (SUBV-2021-168), and PROMPT (#164_Voix-EERS 2021.06), for the ÉTS-EERS Industrial Research Chair in In-Ear Technologies (CRITIAS), sponsored by EERS Global Technologies Inc. This study was supported by the German Research Foundation DFG under Germany's Excellence Strategy (EXC 2177/2 - Project ID 390895286) and the DFG Emmy-Noether program (BL 1591/1-1 - Project ID 411333557).

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.

References

Anderson, C. W., and Sijercic, Z. (1996). "Classification of EEG signals from four subjects during five mental tasks," in *Solving Engineering Problems With Neural Networks: Proceedings of the Conference on Engining Applications in Neural Networks (EANN'96)*, 407–414.

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

Aroudi, A., Mirkovic, B., De Vos, M., and Doclo, S. (2016). "Auditory attention decoding with EEG recordings using noisy acoustic reference signals," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 694-698. doi: 10.1109/ICASSP.2016.7471764

Bauch, C. D., and Olsen, W. O. (1990). Comparison of abr amplitudes with tiptrode^{TM} and mastoid electrodes. *Ear Hear.* 11, 463–467. doi: 10.1097/00003446-199012000-00010

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

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., et al. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviat. Space. Environm. Med.* 78, B231–B244.

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

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

Bleichner, M. G., Lundbeck, M., Selisky, M., Minow, F., Jäger, M., Emkes, R., et al. (2015). Exploring miniaturized EEG electrodes for brain-computer interfaces. An EEG you do not see? *Physiolog. Rep.* 3, 00011. doi: 10.14814/phy2.12362

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

Brouwer, A.-M., Hogervorst, M. A., Erp, J. B. F., Heffelaar, T., Zimmerman, P. H., and Robert, O. (2012). Estimating workload using EEG spectral power and ERPs in the n-back task. *J. Neural Eng.* 9, 045008. doi: 10.1088/1741-2560/9/4/045008

Buzsaki, G. (2011). *Rhythms of the Brain, 1st Edn.* Oxford: Oxford University Press.

Casson, A. J., Yates, D. C., Smith, S. J. M., Duncan, J. S., and Rodriguez-Villegas, E. (2010). Wearable electroencephalography. *IEEE Open J. Eng. Med. Biol.* 29, 44–56. doi: 10.1109/MEMB.2010.936545

Chi, Y., Wang, Y.-T., Wang, Y., Maier, C., Jung, T.-P., and Cauwenberdgs, G. (2011). "Dry and noncontact EEG sensors for mobile brain-computer interfaces," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering : A Publication of the IEEE Engineering in Medicine and Biology Society, Vol. 20* (IEEE), 228–235. doi: 10.1109/TNSRE.2011.2174652

Cohen, M. X. (2014). *Analyzing Neural Time Series Data*. Cambridge: The MIT Press. doi: 10.7551/mitpress/9609.001.0001

Combrisson, E., and Jerbi, K. (2015). Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *J. Neurosci. Methods* 250, 126–136. doi: 10.1016/j.jneumeth.2015.01.010

De Vos, M., Gandras, K., and Debener, S. (2014a). Towards a truly mobile auditory brain–computer interface: exploring the P300 to take away. *Int. J. Psychophysiol.* 91, 46–53. doi: 10.1016/j.ijpsycho.2013.08.010

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

Debener, S., Emkes, R., De Vos, M., and Bleichner, M. (2015). Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear. *Sci. Rep.* 5, 00015. doi: 10.1038/srep16743

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

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

Farquhar, J., and Hill, N. J. (2013). Interactions between pre-processing and classification methods for event-related-potential classification. *Neuroinformatics* 11, 175–192. doi: 10.1007/s12021-012-9171-0

Fougnie, D. (2008). "The relationship between attention and working memory," in *New Research on Short-Term Memory, Vol. 1* (Nova Science Publishers, Inc).

Foxe, J. J., and Snyder, A. C. (2011). The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Front. Psychol.* 2, 00242. doi: 10.3389/fpsyg.2011.00154

Goverdovsky, V., von Rosenberg, W., Nakamura, T., Looney, D., Sharp, D. J., Papavassiliou, C., et al. (2017). Hearables: multimodal physiological in-ear sensing. *Sci. Rep.* 7, 6948. doi: 10.1038/s41598-017-06925-2

Gramann, K., Jung, T.-P., Ferris, D. P., Lin, C.-T., and Makeig, S. (2014). Toward a new cognitive neuroscience: modeling natural brain dynamics. *Front. Hum. Neurosci.* 8, 7. doi: 10.3389/978-2-88919-271-7

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2014). MNE software for processing MEG and EEG data. *Neuroimage* 86, 446–460. doi: 10.1016/j.neuroimage.2013.10.027

Harmony, T., Fernández, T., Silva, J., Bernal, J., Díáz-Comas, L., Reyes, A., et al. (1996). EEG delta activity: an indicator of attention to internal processing during performance of mental tasks. *Int. J. Psychophysiol.* 24, 161–171. doi: 10.1016/S0167-8760(96)00053-0

Herrmann, C. S., Munk, M. H. J., and Engel, A. K. (2004). Cognitive functions of gamma-band activity: memory match and utilization. *Trends Cogn. Sci.* 8, 347–355. doi: 10.1016/j.tics.2004.06.006

Hochberg, L. R., Serruya, M. D., Friehs, G. M., Mukand, J. A., Saleh, M., Caplan, A. H., et al. (2006). Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature* 442, 164–171. doi: 10.1038/nature04970

Jarchum, I. (2019). The ethics of neurotechnology. *Nat. Biotechnol.* 37, 993–996. doi: 10.1038/s41587-019-0239-3

Jenkins, J. L., Anderson, B. B., Vance, A., Kirwan, C. B., and Eargle, D. (2016). More harm than good? How messages that interrupt can make us vulnerable. *Inform. Syst. Res.* 27, 880–896. doi: 10.1287/isre.2016.0644

Keil, A., Mussweiler, T., and Epstude, K. (2006). Alpha-band activity reflects reduction of mental effort in a comparison task: a source space analysis. *Brain Res.* 1121, 117–127. doi: 10.1016/j.brainres.2006.08.118

Kidmose, P., Looney, D., and Mandic, D. P. (2012). "Auditory evoked responses from Ear-EEG recordings," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 586–589. doi: 10.1109/EMBC.2012.6345999

Klimesch, W., Doppelmayr, M., Russegger, H., and Pachinger, T. (1996). Theta band power in the human scalp EEG and the encoding of new information. *Neuroreport* 7, 1235–1240. doi: 10.1097/00001756-199605170-00002

Klimesch, W., Doppelmayr, M., Russegger, H., Pachinger, T., and Schwaiger, J. (1998). Induced alpha band power changes in the human EEG and attention. *Neurosci. Lett.* 244, 73–76. doi: 10.1016/S0304-3940(98)00122-0

Kothe, C. (2014). *Lab Streaming Layer (LSL)*. Available online at: https://github.com/sccn/labstreaminglayer (accessed October 26, 2015).

Krusienski, D. J., Sellers, E. W., Cabestaing, F., Bayoudh, S., Mcfarland, D. J., Vaughan, T. M., et al. (2006). A comparison of classification techniques for the P300 speller. *J. Neural Eng.* 3, 299–305. doi: 10.1088/1741-2560/3/4/007

Landau, A. N., Esterman, M., Robertson, L. C., Bentin, S., and Prinzmetal, W. (2007). Different effects of voluntary and involuntary attention on EEG activity in the gamma band. *J. Neurosci.* 27, 11986–11990. doi: 10.1523/JNEUROSCI.3092-07.2007

Lin, C.-T., Chuang, C.-H., Huang, C.-S., Tsai, S.-F., Lu, S.-W., Chen, Y.-H., et al. (2014). Wireless and wearable EEG system for evaluating driver vigilance. *IEEE Trans. Biomed. Circuits Syst.* 8, 165–176. doi: 10.1109/TBCAS.2014.2316224

Looney, D., Kidmose, P., Morrell, M. J., and Mandic, D. P. (2014). "Ear-EEG: continuous brain monitoring," in *Brain-Computer Interface Research, SpringerBriefs in Electrical and Computer Engineering*. Cham: Springer International Publishing, 63–71. doi: 10.1007/978-3-319-09979-8_6

Looney, D., Park, C., Kidmose, P., Rank, M., Ungstrup, M., Rosenkranz, K., et al. (2011). "An in-the-ear platform for recording electroencephalogram," in *2011 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Boston, MA: IEEE, 6882–5. doi: 10.1109/IEMBS.2011.6091733

Magosso, E., De Crescenzio, F., Ricci, G., Piastra, S., and Ursino, M. (2019). EEG alpha power is modulated by attentional changes during cognitive tasks and virtual reality immersion. *Comput. Intell. Neurosci.* 2019, 7051079. doi: 10.1155/2019/7051079

Mahjoory, K., Nikulin, V. V., Botrel, L., Linkenkaer-Hansen, K., Fato, M. M., and Haufe, S. (2017). Consistency of EEG source localization and connectivity estimates. *Neuroimage* 152, 590–601. doi: 10.1016/j.neuroimage.2017.02.076

Mijović, P., Ković, V., Vos, M. D., Mačužić, I., Todorović, P., Jeremić, B., et al. (2017). Towards continuous and real-time attention monitoring at work: reaction time versus brain response. *Ergonomics* 60, 241–254. doi: 10.1080/00140139.2016.1142121

Mikkelsen, K. B., Ebajemito, J. K., Bonmati-Carrion, M. A., Santhi, N., Revell, V. L., Atzori, G., et al. (2018). Machine-learning-derived sleep–wake staging from around-the-ear electroencephalogram outperforms manual scoring and actigraphy. *J. Sleep Res.* 0, e12786. doi: 10.1111/jsr.12786

Mikkelsen, K. B., Kappel, S. L., Mandic, D. P., and Kidmose, P. (2015). EEG recorded from the ear: characterizing the ear-EEG method. *Front. Neurosci.* 9, 00003. doi: 10.3389/fnins.2015.00438

Mirkovic, B., Bleichner, M. G., De Vos, M., and Debener, S. (2016). Target speaker detection with concealed EEG around the ear. *Auditory Cognit. Neurosci.* 2016, 349. doi: 10.3389/fnins.2016.00349

Müller, K. R., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., and Blankertz, B. (2008). Machine learning for real-time single-trial EEG-analysis: From brain–computer interfacing to mental state monitoring. *J. Neurosci. Methods.* 167, 82–90. doi: 10.1016/j.jneumeth.2007.09.022

Oken, B. S., Salinsky, M. C., and Elsas, S. M. (2006). Vigilance, alertness, or sustained attention: physiological basis and measurement. *Clini. Neurophysiol.* 117, 1885–1901. doi: 10.1016/j.clinph.2006.01.017

Owen, A. M., McMillan, K. M., Laird, A. R., and Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapp.* 25, 46-59. doi: 10.1002/hbm.20131

Pacharra, M., Debener, S., and Wascher, E. (2017). Concealed around-the-ear EEG captures cognitive processing in a visual simon task. *Front. Hum. Neurosci.* 11, 290. doi: 10.3389/fnhum.2017.00290

Palaniappan, R. (2006). Utilizing gamma band to improve mental task based brain-computer interface design. *IEEE Trans. Neural Syst. Rehabil.* 14, 299–303. doi: 10.1109/TNSRE.2006.881539

Pesonen, M., Hämäläinen, H., and Krause, C. M. (2007). Brain oscillatory 4–30 Hz responses during a visual n-back memory task with varying memory load. *Brain Res.* 1138, 171–177. 00000. doi: 10.1016/j.brainres.2006.12.076

Pfotenhauer, S. M., Frahm, N., Winickoff, D., Benrimoh, D., Illes, J., and Marchant, G. (2021). Mobilizing the private sector for responsible innovation in neurotechnology. *Nat. Biotechnol.* 39, 661–664. doi: 10.1038/s41587-021-00947-y

Popescu, F., Fazli, S., Badower, Y., Blankertz, B., and Müller, K.-R. (2007). Single Trial Classification of Motor Imagination Using 6 Dry EEG Electrodes. *PLoS ONE* 2(7):e637. doi: 10.1371/journal.pone.0000637

Rebsamen, B., Kwok, K., and Penney, T. B. (2011). Evaluation of cognitive workload from EEG during a mental arithmetic task. *Proc. Hum. Factors Ergon.* 55, 1342–1345. doi: 10.1177/1071181311551279

Regenbogen, C., Vos, M. D., Debener, S., Turetsky, B. I., Mößnang, C., Finkelmeyer, A., et al. (2012). Auditory processing under cross-modal visual

load investigated with simultaneous EEG-fMRI. *PLoS ONE*. 7, e52267.e00000. doi: 10.1371/journal.pone.0052267

Scheeringa, R., Bastiaansen, M. C. M., Petersson, K. M., Oostenveld, R., Norris, D. G., and Hagoort, P. (2008). Frontal theta EEG activity correlates negatively with the default mode network in resting state. *Int. J. Psychophysiol*. 67, 242–251. doi: 10.1016/j.ijpsycho.2007.05.017

Shibata, T., Shimoyama, I., Ito, T., Abla, D., Iwasa, H., Koseki, K., et al. (1999). Attention changes the peak latency of the visual gamma-band oscillation of the EEG. *Neuroreport* 10, 1167–1170. doi: 10.1097/00001756-199904260-00002

Sterr, A., Ebajemito, J. K., Mikkelsen, K. B., Bonmati-Carrion, M. A., Santhi, N., della Monica, C., et al. (2018). Sleep EEG derived from behind-the-ear electrodes (cEEGrid) compared to standard polysomnography: a proof of concept study. *Front. Hum. Neurosci*. 12, 452. doi: 10.3389/fnhum.2018.00452

Trujillo, L. T., Peterson, M. A., Kaszniak, A. W., and Allen, J. J. B. (2005). EEG phase synchrony differences across visual perception conditions may depend on recording and analysis methods. *Clini. Neurophysiol*. 116, 172–189. doi: 10.1016/j.clinph.2004.07.025

Uhlhaas, P. J., Pipa, G., Neuenschwander, S., Wibral, M., and Singer, W. (2011). A new look at gamma? High- (>60 Hz) -band activity in cortical networks: function, mechanisms and impairment. *Prog. Biophys. Mol*. 105, 14–28. doi: 10.1016/j.pbiomolbio.2010.10.004

Vansteensel, M. J., Pels, E. G., Bleichner, M. G., Branco, M. P., Denison, T., Freudenburg, Z. V., et al. (2016). Fully implanted brain–computer interface in a locked-in patient with ALS. *N. Engl. J. Med*. 375, 2060–2066. doi: 10.1056/NEJMoa1608085

Watter, S., Geffen, G. M., and Geffen, L. B. (2001). The n-back as a dual-task: P300 morphology under divided attention. *Psychophysiology* 38, 998–1003. doi: 10.1111/1469-8986.3860998

Wexler, A. (2021). A skeptic's take on Neuralink and other consumer neurotech. *Stat News*. Available online at: https://www.statnews.com/2021/04/07/consumer-neurotech-skeptic/

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

Yekutieli, D., and Benjamini, Y. (1999). Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *J. Stat. Plan. Inference* 82, 171–196. doi: 10.1016/S0378-3758(99)00041-5

Yu, X., Zhang, J., Xie, D., Wang, J., and Zhang, C. (2009). Relationship between scalp potential and autonomic nervous activity during a mental arithmetic task. *Auto. Neurosci*. 146, 81–86. doi: 10.1016/j.autneu.2008.12.005

Yuval-Greenberg, S., Tomer, O., Keren, A. S., Nelken, I., and Deouell, L. Y. (2008). Transient induced gamma-band response in EEG as a manifestation of miniature saccades. *Neuron* 58, 429–441. doi: 10.1016/j.neuron.2008.03.027

Zander, T. O., and Kothe, C. (2011). Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. *J. Neural Eng*. 8, 025005.000307. doi: 10.1088/1741-2560/8/2/025005

Zivkovic, M. (2015). Brain culture: neuroscience and popular media. *Interdiscipl. Sci. Rev*. 40, 409–417. doi: 10.1080/03080188.2016.1165457

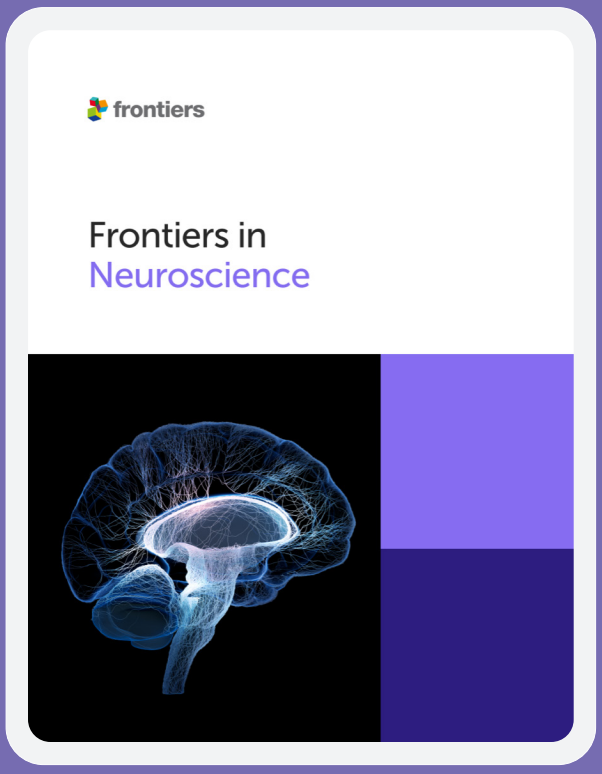
Frontiers in Neuroscience

Provides a holistic understanding of brain function from genes to behavior

Part of the most cited neuroscience journal series which explores the brain - from the new eras of causation and anatomical neurosciences to neuroeconomics and neuroenergetics.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact