Novel AIRTrode-based wearable electrode supports long-term, online brain–computer interface operations

Journal of Neural Engineering

PAPER

Novel AIRTrode-based wearable electrode supports long-term, online brain–computer interface operations

Deland H Liu[1,6,*], **Ju-Chun Hsieh**[2,6,*], **Hussein Alawieh**[1], **Satyam Kumar**[1], **Fumiaki Iwane**[5], **Ilya Pyatnitskiy**[2], **Zoya J Ahmad**[2], **Huiliang Wang**[2,7,*] and **José del R Millán**[1,2,3,4,7,*]

[1] Chandra Department of Electrical and Computer Engineering, Cockrell School of Engineering, The University of Texas at Austin, Austin 78712 TX, United States of America
[2] Department of Biomedical Engineering, Cockrell School of Engineering, The University of Texas at Austin, Austin 78712 TX, United States of America
[3] Department of Neurology, Dell Medical School, The University of Texas at Austin, Austin 78712 TX, United States of America
[4] Mulva Clinic for the Neurosciences, The University of Texas at Austin, Austin 78712 TX, United States of America
[5] National Institute of Neurological Disorders and Stroke, National Institute of Health, Bethesda 20892 MD, United States of America
[6] These authors contributed equally to the work.
[7] These authors jointly supervised this work.
[*] Authors to whom any correspondence should be addressed.

E-mail: deland.liu@utexas.edu, juchun.hsieh@utexas.edu, jose.millan@austin.utexas.edu and evanwang@utexas.edu

Keywords: brain–computer interface (BCI), hydrogel electrode, electroencephalogram (EEG), motor imagery (MI), error-related potential (ErrP), wearable technologies

Supplementary material for this article is available online

OPEN ACCESS

RECEIVED
30 June 2024

REVISED
18 September 2024

ACCEPTED FOR PUBLICATION
13 December 2024

PUBLISHED
7 January 2025

Abstract

Objective. Non-invasive electroencephalograms (EEG)-based brain–computer interfaces (BCIs) play a crucial role in a diverse range of applications, including motor rehabilitation, assistive and communication technologies, holding potential promise to benefit users across various clinical spectrums. Effective integration of these applications into daily life requires systems that provide stable and reliable BCI control for extended periods. Our prior research introduced the AIRTrode, a self-adhesive (A), injectable (I), and room-temperature (RT) spontaneously-crosslinked hydrogel electrode (AIRTrode). The AIRTrode has shown lower skin-contact impedance and greater stability than dry electrodes and, unlike wet gel electrodes, does not dry out after just a few hours, enhancing its suitability for long-term application. This study aims to demonstrate the efficacy of AIRTrodes in facilitating reliable, stable and long-term online EEG-based BCI operations. *Approach.* In this study, four healthy participants utilized AIRTrodes in two BCI control tasks–continuous and discrete–across two sessions separated by six hours. Throughout this duration, the AIRTrodes remained attached to the participants' heads. In the continuous task, participants controlled the BCI through decoding of upper-limb motor imagery (MI). In the discrete task, the control was based on decoding of error-related potentials (ErrPs). *Main Results.* Using AIRTrodes, participants demonstrated consistently reliable online BCI performance across both sessions and tasks. The physiological signals captured during MI and ErrPs tasks were valid and remained stable over sessions. Lastly, both the BCI performances and physiological signals captured were comparable with those from freshly applied, research-grade wet gel electrodes, the latter requiring inconvenient re-application at the start of the second session. *Significance.* AIRTrodes show great potential promise for integrating non-invasive BCIs into everyday settings due to their ability to support consistent BCI performances over extended periods. This technology could significantly enhance the usability of BCIs in real-world applications, facilitating continuous, all-day functionality that was previously challenging with existing electrode technologies.

1. Introduction

In recent years, brain–computer interfaces (BCIs) have emerged as promising communication and assistive tools for individuals with disabilities, offering them the ability to interact with their surroundings and control external devices [1]. Among various noninvasive BCI methods, substantial progress has been achieved in electroencephalogram (EEG)-based BCIs due to EEG's high temporal resolution, cost-effectiveness, and the overall ease of use of EEG systems [2, 3]. Two particular EEG-BCIs that have achieved success are motor imagery (MI) and error-related potential (ErrP)-based BCIs. MI involves the mental simulation of limb movement [4], producing distinct EEG patterns that BCIs can decode to operate assistive technologies such as wheelchairs [5], robotic prostheses [6, 7], telepresence robots [8], and spelling systems [9]. Different from MI, ErrPs are event-related potentials (ERPs) elicited upon the perception of an error [10]. In prior research, ErrP-based BCIs have been integrated in spellers to facilitate efficient communication in clinical populations such as those with amyotrophic lateral sclerosis (ALS) [11] and have been used to personalize robot trajectories for obstacle avoidance [12].

Despite these advancements, the practical application of EEG-based BCIs requires long-term, stable signal acquisition, a significant challenge with traditional gel and dry electrodes. The gold-standard in research laboratories for EEG recording is the gel silver/silver chloride (Ag/AgCl) electrodes. These electrodes rely on conductive gel to reduce the impedance between the skin and the electrodes, thereby minimizing noise from interface mismatch. The effectiveness of gel electrodes largely stems from their ability to conform to the skin, yielding low impedance and high-quality EEG signals [13]. However, their major drawback is that the conductive gel typically dries out within about 4 h, leading to deterioration of impedance and rendering them unsuitable for prolonged EEG monitoring [14].

In contrast, dry electrodes, made from solid conductive materials such as metals, conductive textiles, and polymers [15, 16], do not dry out, offer simpler usage and greater accessibility for untrained users [17, 18]. Yet, they lack the necessary softness and adhesiveness to establish robust contact with the scalp, particularly through hair. To mitigate this, surface structural designs like microneedle or pillar-shaped configurations are often utilized [16, 19, 20]. However, these designs can require external forces to maintain contact, causing discomfort during prolonged use [16, 21, 22]. Furthermore, the long-term use of dry electrodes is challenged by reduced impedance stability due to insufficient contact, as well as deterioration caused by ambient humidity and sweat-induced corrosion. These factors increase susceptibility to motion artifacts and degrade signal quality over time [23].A recent development in dry electrodes is the carbon nanotube (CNT)/adhesive polydimethylsiloxane (aPDMS) composite-based electrode, which has shown good adhesion, contact and signal-to-noise ratio (SNR) in EEG recordings [24]. Similar to some hydrogel electrodes, this electrode addresses the challenges of recording EEG signals through hair, achieving comparable signal quality to conventional gel electrodes. While these electrodes show promise for long-term stability and reliable BCI performance, their long-term reliability has yet to be validated.

The shortcomings of conventional gel and dry electrodes in long-term use have prompted researchers to explore alternatives. One such example is the MXtrodes–a non-hydrogel, bioelectronic interface built from MXenes–transition metal carbides, nitrides, and carbonitrides. Unlike gel electrodes, MXtrodes do not require wet gels, allowing them to maintain stable electrode-skin interface impedance and recording quality over time without drying out. Leveraging the unique properties of MXenes, MXtrodes offer low-cost, rapid, and scalable processing [25–27]. With these properties, MXtrodes show great potential for long-term, widespread, high-precision recordings, though their long-term stability has yet to be fully demonstrated in the literature. Another promising alternative is conductive hydrogel electrodes, which utilize self-electrolytes to form ion channels at the skin interface, eliminating the need for conductive gel [28]. As a result, hydrogel electrodes support the acquisition of high-quality EEG signals over extended periods, effectively addressing the limitations of gel electrodes. Additionally, the soft and adhesive properties of hydrogels allow these electrodes to penetrate hair and establish robust contact with the scalp without the need for needle-shaped designs or the application of external force, overcoming the primary issues associated with dry electrodes. Existing literature has shown the long-term impedance stability of hydrogel electrodes [28–31] and their efficacy in passive, long-term EEG recordings [23, 32, 33].

Recently, we introduced AIRTrode, an on-site formed self-adhesive (A), injectable (I), and room-temperature (RT) spontaneously-crosslinked hydrogel electrode (AIRTrode), designed for long-term sleep EEG monitoring applications [34]. In this previous work, we showcased AIRTrode's ability to maintain low impedance of 17.53 $k\Omega$ over eight hours of continuous wear on hairy scalp regions. Additionally, these electrodes showed a high adhesiveness of 0.92 N/cm^2 and repeatable adhesion, confirming

their suitability for long-term usage during both daily activities and overnight sleep. Moreover, AIRTrode exhibited a superior signal-to-noise ratio (SNR of 23.97 *dB*), significantly outperforming commercial gel electrodes (SNR of 17.98 *dB*), and matched the performance of research-grade gel electrodes in sleep stage classification [34]. For a visual illustration of the AIRTrode electrode, please refer to supplementary figure S1

Despite recent advancements in conductive hydrogel electrodes and existing EEG studies investigating long-term recordings over periods of five-six days [34, 35], to our knowledge, no existing studies have demonstrated a hydrogel-based electrode capable of supporting reliable, online EEG-based BCI operations for extended periods. Building on the demonstrated long-term stability and efficacy of AIRTrode in EEG monitoring, we hypothesize that AIRTrode can robustly support reliable MI and ErrP-based BCI operations for periods exceeding six hours without degrading the BCIs' online performance or the quality of the captured physiological patterns. Moreover, we expect that BCI performance with AIRTrode will be statistically indistinguishable, if not exceed, the benchmarks set by traditional, research-grade gel electrodes, ensuring no significant deviation in effectiveness across extended usage.

To investigate these hypotheses, we carried out a four-day cross-over experiment, with each day spaced a week apart. On days one and two, participants used AIRTrode and gel electrodes, respectively, to control a left hand (LH) versus right hand (RH) MI-based BCI. Days three and four followed the same pattern but for an ErrP-based BCI. Each experimental day consisted of morning and afternoon sessions, separated by a six-hour break during which participants continued with their usual activities, such as attending lectures, eating lunch, and riding bicycles. For each electrode type, we assessed the stability of neurophysiological properties and online performance across morning and afternoon sessions for each BCI paradigm, and compared them to the benchmarks established by research-grade gel electrodes.

2. Materials and methods

2.1. Participants

We recruited four participants (3 self-reported Asian males and 1 self-reported male of Arab ethnicity, age range of 25-30 years, M = 26.75, SD = 2.06), all right-handed. Participants were designated as 'male' based on their self-reported sex. The participants had diverse hair length, texture and color. Two participants had medium-length, dense black hair, approximate 10–12 cm in length. One participant had medium-length, fine black hair, approximately 6–8 cm in length. Another participant had long, curly, dense dark brown hair, approximately 15 cm in length. Inclusion criteria included normal or corrected-to-normal vision, absence of neurological disorders, and no use of psychoactive medications. All participants provided written informed consent in accordance with the Declaration of Helsinki. The study protocol was approved by the local ethics committee (2020–03-0073, The University of Texas at Austin, TX, USA).

2.2. Experimental paradigm and set-up

Subjects participated in a 4-days cross-over study. On each day, participants engaged in either MI-based or ErrP-based BCI control, using either AIRTrode or gel electrodes. Days were divided into morning and afternoon sessions, each lasting approximately 90 min, separated by a six-hour interval (figure 1). In each recording session, the participant sat on a comfortable chair in front of a computer screen (14-inch display, 2560 × 1440 pixels, 60 Hz refresh rate, ThinkPad X1 Carbon), which displayed either the MI or ErrP task.

Morning sessions began with a calibration phase to train personalized decoders, followed by an online BCI operation phase using these decoders. Afternoon sessions resumed with participants returning to use the same decoders for further online BCI operation.

In days of MI-based BCI recordings, during offline sessions, participants completed four runs of MI tasks involving pseudo-random cues for LH or RH movements. Each run included 10 trials per class, used to construct a binary-class BCI. Online sessions included three runs following the same structure, with participants receiving real-time feedback on their performance.

In days of ErrP-based BCI recordings, offline sessions comprised three runs where participants passively observed actions performed by an external agent, classified as correct or erroneous (100 trials per run). Online sessions followed the same structure and included performance feedback. An additional offline session was recorded before the afternoon session each day for neurophysiological analysis purposes, not for decoder training.

For AIRTrode recordings, AIRTrode electrodes were kept in place throughout the day, including during the six-hour inter-session break. Gel electrodes, however, were removed and reapplied fresh for each afternoon session.

2.3. AIRTrode

The AIRTrode hydrogel electrodes were synthesized according to the previous report with no modifications [34]. In brief, Dimethyl sulfoxide (DMSO) was first blended with poly (3, 4-ethylenedioxythiophene) polystyrene sulfonate (PEDOT:PSS) at a 4.5% weight ratio relative to

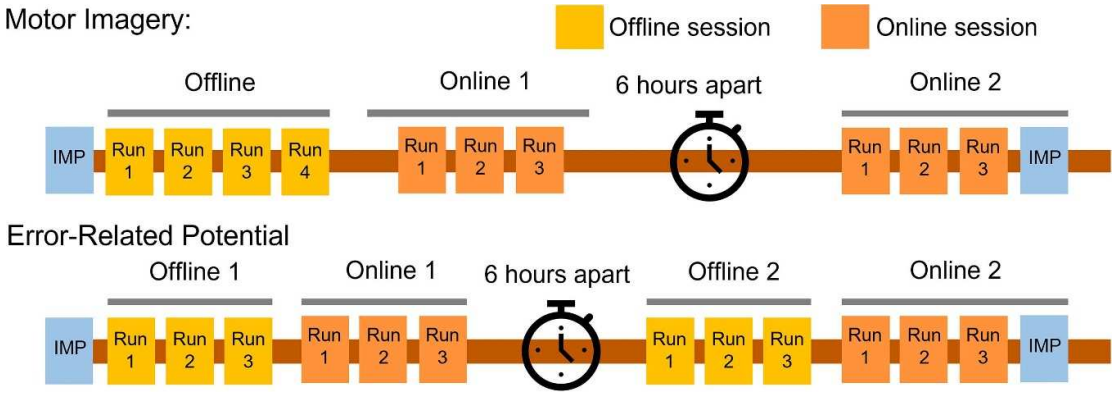

Figure 1. Study protocol. Each recording day involved participants undergoing MI or ErrP-based BCI control with either AIRTrode or research-grade gel electrodes. Impedance measurements (IMP) were taken at the start and end of each recording day. Subject-specific decoders were trained after the morning offline session and used in the following two online sessions. AIRTrode remained in place during the six-hour break while participants engaged in daily activities; gel electrodes were reapplied post-break.

PEDOT:PSS. A vortex mixer (Corning LSE Vortex Mixer, Fisher Scientific) was used for 30 s to prepare a homogeneous mixture. This mixture was then further blended with 17.3% weight ratio glycerol relative to PEDOT:PSS for 30 s using a vortex mixer. Next, 2-Acrylamido-2-methylpropane sulfonic acid (AMPS) monomer with 73.5% weight ratio relative to PEDOT:PSS was added to the mixture in the same vortex mixer and mixed for 60 s to obtain an uniform dispersion. The final mixture was then loaded to a syringe, and placed at room temperature. AIRTrode was spontaneously formed without any external stimulus or crosslinking reagent within 30 min.

2.4. EEG acquisition

EEG signals were recorded from 13 electrodes: Fz, FC5, FC1, FCz, FC2, FC6, C3, Cz, C4, CP5, CP1, CP2 and CP6, according to the 10/20 international system. GND was placed at AFz and reference at CPz. EEG was recorded at 512 Hz using an eego system (ANT Neuro, Netherlands). The hydrogel electrodes were applied via a syringe to a waveguard EEG cap (ANT Neuro, Netherlands), which was mounted on the participant's head. Electrode impedance was acquired using eego software (ANT Neuro, Netherlands).

2.5. MI experiments

2.5.1. Brain–computer interface task

The 'bar feedback task' used in the MI experiment is based on traditional protocols that have been commonly used in the BCI research field over the past decade. The 'bar feedback task' or its slight variations have been employed in several prior studies [36–42]. In this task, participants received visual feedback through a bar displayed on the screen, extending from one side to the other, with the left-hand side indicating the LH class and the right-hand side the RH class. Each trial started with a 2 s rest period (± 1 s), followed by a 2 s fixation period. Subsequently, a task cue for 2 s indicated the trial's class: a blue arrow pointing left for LH and a red arrow pointing right for RH. In offline runs, the task execution period lasted 5 s, with a simulated BCI output moving the bar towards the indicated direction, providing continuous visual feedback. The participant performed MI, as directed by this feedback. During online runs, the BCI output controlled the direction of the bar. Specifically, the bar moved in proportion to the accumulated evidence from the BCI decoder for the two MI classes. Two decision thresholds were represented by lines at opposite ends of the screen. If the bar reached a decision threshold, discrete feedback in the form of an arrow appeared for 2 s to indicate the executed command. In online runs, the feedback duration was capped at 7 s. If a decision threshold was not reached within this time, the trial ended with a 'Timeout'.

To perform MI correctly, subjects were given the following instructions: 'Based on the presented task cue, you have to mentally rehearse the kinesthetic experience for a single sustained movement of your right or left hand. You must not physically perform the movement nor elicit any muscular contraction. Make sure to attempt consistent imagination of the movements over sessions.'

2.5.2. MI classification

During online operation, EEG signals were bandpass filtered within the [8,30] Hz range using a second-order Butterworth filter. This frequency range was selected because MI is typically associated with event-related desynchronizations (ERDs) in the sensorimotor rhythms (SMRs), particularly within the μ

([8,13] Hz) and β ([18,30] Hz) bands. By choosing this specific frequency range, we aimed to effectively capture the neural patterns specific to MI and reduce noise from irrelevant bands, thereby improving the performance of the BCI [43–46]. Features were extracted from 1-second sliding windows, updated every 62.5 ms, by calculating the covariance matrices of the EEG signals within these windows. A Riemannian geometry based-decoder was used, utilizing a minimum-distance-to-mean classification scheme on the Riemannian manifold comprised of spatial covariance matrices. This classifier operates similarly to a nearest neighbor classifier, categorizing covariance matrices into distinct classes based on their distance from class-specific means located on the Riemannian manifold [47]. We chose to use a classification method based on Riemannian geometry for several reasons. First, Riemannian geometry-based methods have demonstrated accurate and robust classification in EEG MI-based BCIs in prior literature [48, 49], especially in online experiments [42, 50]. During the pilot stage of our experiment, we initially used a linear algorithm, LDA, but it underperformed in both offline and online performances compared to Riemannian methods, especially given the limited number of channels available. Second, in prior literature, Riemannian geometry-based methods have proven to be robust in capturing non-stationarities in EEG signals over multiple sessions [42, 50–52]. This robustness is particularly relevant in our study, where EEG signals recorded over a 6-hour timeframe can exhibit non-stationary behavior, with statistical properties changing over time. By working on the Riemannian manifold of covariance matrices, this approach enables more reliable tracking of underlying signal changes, which is crucial for maintaining performance in a long-duration MI BCI. For further details on the classifier's methodology and implementation, refer to [47, 50].

To deliver a command (RH versus LH), the BCI accumulates estimated probabilities for each class until one class's accumulated probability reaches a pre-set threshold. A trial is deemed a 'hit' if the accumulated probability meets the threshold for the correct class, and a 'miss' if it reaches the threshold for the incorrect class. If neither threshold is reached within the designated timeout period, the trial is classified as a 'Timeout'. Evidence for each class is accumulated using an exponential smoothing formula [40]:

$$e_i = 0.95 \cdot e_{i-1} + 0.05 \cdot p_i \quad (1)$$

where e_i is the accumulated evidence for a given class and p_i is the posterior probability of that class for the i^{th} sample. Evidence accumulation is reset to 0.5 at the beginning of every trial. The thresholds were adjusted to provide a suitable challenge level while minimizing the likelihood of misses.

2.5.3. BCI performance analysis

To assess online BCI performance, we evaluated the accuracy of BCI output using sample-wise classification accuracy, user's control level using the bar dynamics metric, and command delivery performance using the BCI hits accuracy metric.

- Sample-wise classification accuracy: this measures the percentage of samples correctly classified as the respective class (RH vs LH) in each run

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}, \quad (2)$$

  where *TP*, *TN*, *FP* and *FN* correspond respectively to true positives, true negatives, false positives and false negatives. Contrarily to trial-level performance, sample-level accuracy is independent of the thresholds used for evidence accumulation.
- Bar dynamics: this reflects the extent of BCI control by indicating the percentage of time during the MI task execution period of a trial that the accumulated evidence from the BCI decoder supported the correct class. It was calculated based on the proportion of samples for which the accumulated probability for the correct class exceeded 50%.
- BCI hits accuracy: this metric calculates the ratio of trials with correct threshold hits to the total number of threshold crossings within a session, excluding trials with a 'Timeout' outcome. The metric was then normalized by $1 - \frac{n_{\text{timeouts}}}{n_{\text{trials}}}$, where n_{timeouts} is the number of trials that concluded with 'Timeout' and n_{trials} is the total number of trials in a session.

The chance levels for the classification accuracy and the command delivery metrics above were computed by randomly permuting the testing labels 10 000 times and averaging the corresponding outcomes over the permutations.

2.5.4. MI neurophysiology analysis.

To compute event-related desynchronizations (ERDs) in the sensorimotor rhythms (SMRs) within the μ ([8,13] Hz) and β ([18,30] Hz) bands, EEG signals underwent several processing steps. Initially, signals were spatially filtered using a surface Laplacian to enhance spatial resolution. They were then band-pass filtered within the specified frequency ranges using a 4th order non-causal Butterworth filter. Next, we calculated the band-limited power of the filtered signals by averaging the squared voltage amplitudes for each electrode during the task execution period of each

trial. This task-relevant band power was then normalized by subtracting and dividing by the baseline power, which was computed from the pre-task fixation period [44]. ERDs were averaged over all trials and participants. For visualization, ERDs from right hand (RH) trials were medially transposed, ensuring that contra-lateral ERDs consistently appeared on the right hemisphere in topological plots.

2.6. ErrP experiments

2.6.1. Brain–computer interface task

We used a cursor task in the ErrP experiments, based on established protocols widely employed in BCI research over the past decade. This task, along with its variations, has been utilized in several previous studies [53–57]. Similar to before, participants monitored a dynamic cursor, represented by a blue circle ($3.14, \mathrm{cm}^{-2}$), moving across 10 square locations ($4\,\mathrm{cm}^{-2}$) along the screen's central horizontal plane. The cursor's target square location was highlighted in red. At each time step, termed a trial here, the cursor moved horizontally to an adjacent square location, either to the right or left, ensuring it remained within the predefined working area. Each trial lasted between 2000 and 2500 ms. The cursor movement stopped once the target was reached, after which a new cursor and target were randomly placed among the 10 squares.

Participants were instructed to fix their gaze on the center of the screen and passively monitor the cursor's movement, expecting it to move in the direction towards the target. An ErrP was elicited whenever the cursor moved away from the target, contrary to the participants' expectation. Each run consisted of 100 trials lasting about 3 min, with a 30% error probability per trial where the cursor moved in the erroneous direction (i.e. direction contrary to the participants' expectation).

Our protocol included BCI feedback during the online operation of the ErrP task. Specifically, if the BCI falsely detected the presence of an ErrP following a correct cursor movement, the cursor would jump two locations further from the target in the next trial. If this displacement moved the cursor outside the working area, it would instead move by only one location. Conversely, if the BCI correctly detected the presence of an ErrP following an erroneous cursor movement, the cursor would move two locations closer to the target on the next trial. The primary goal during online sessions was to move the cursor to the target location as many times as possible, encouraging participants to elicit accurate neural responses—i.e. to generate an ErrP for erroneous actions and not elicit an ErrP for correct actions. Similar to the offline run, the online run consisted of 100 trials lasting about 3 min, with a 30% error probability per trial.

2.6.2. ErrP classification

To build the decoder to detect presence/absence of ErrPs in each online trial, EEG signals were first preprocessed with a 4th order causal Butterworth bandpass filter within the [1, 10] Hz range. These filtered signals were then segmented into epochs using window of [0.20, 0.60] s post-cursor onset. This specific window was chosen to cover the critical error-related negativity (ERN) and error-positivity (Pe) components of the ErrP, as highlighted in the grand average (figure 5). Epochs containing any EEG sample that exceeded 50 μV on any channel were excluded due to artifacts. This exclusion criterion led to the identification of one artefactual epoch in total across the AIRTrode recordings, with no artefactual epochs identified in the research-grade gel recordings, spanning four subjects and two sessions.

To enhance the signal-to-noise ratio (SNR), a spatial filter using canonical correlation analysis (CCA) was applied to the offline epochs. This CCA filter transformed the EEG data to a subspace where the correlation between the epochs and the grand average ErrPs was maximized [58]. We retained the first three CCA components. On each CCA component, we extracted two types of features: (1) The EEG voltage for each time sample after downsampling to 32 Hz (60 temporal features), and (2) Power spectral density (PSD) computed at [4, 6, 8, 10] Hz frequency bins using the Welch method (12 PSD features) [59]. These 72 features per epoch were normalized to [0, 1] to form the feature vector x [12]. From the feature vector x, the ErrP decoder estimated the posterior probability about the presence of an ErrP, $p(\mathrm{error}|x)$, using diagonal linear discriminant analysis:

$$p\left(\mathrm{error}|x\right) = \frac{1}{1-\exp^{-(w'x+b)}} \tag{3}$$

where w represents the weights applied to the input features and b is the bias term. An epoch was classified as containing an ErrP if $p(error|x)$ exceeded a predetermined threshold, τ. This threshold τ was optimized through a leave-one-run-out cross-validation on the offline data [60]. We evaluated the average receiver operating curve (ROC) for various τ values ranging from 0 to 1 in steps of 0.025. A true positive (TP) was defined as an epoch where an ErrP was correctly detected, and a false positive (FP) was an epoch where an ErrP was incorrectly flagged. The optimal τ was chosen to maximize the product of the average true positive rate (TPR) and true negative rate (TNR) from the average ROC during the cross-validation process, and was used in subsequent online sessions.

2.6.3. BCI performance analysis

To evaluate the performance of the ErrP-based BCI performance, we used the following metrics:

- Kappa value: Due to the imbalanced nature of the number of correct and error trials, we computed Cohen's Kappa value to evaluate the classification performance on the imbalanced data, following [61]:

$$\text{Kappa} = \frac{p_a - p_e}{1 - p_e} \tag{4}$$

  where p_a is the classification accuracy and p_e is the probability of a chance detection within the provided dataset.
- True Negative Rate (TNR): This is defined as the proportion of error trials in which the presence of an ErrP was correctly detected (true negatives). TNR is calculated by dividing the number of true negative trials by the total number of error trials.
- True Positive Rate (TPR): This is defined as the proportion of correct trials in which no ErrP was correctly detected (true positives). TPR is calculated by dividing the number of true positive trials by the total number of correct trials.

To calculate the chance levels for Kappa value, TNR and TPR, we randomly permuted the testing labels 10 000 times and averaged the corresponding outcomes over the permutations.

2.6.4. ErrP neurophysiological analyses

Neurophysiological analyses of ErrPs, specifically the analyses of ERN and Pe amplitudes and latencies, peak-to-peak (pk-to-pk) amplitudes of the ErrP and individual θ peak frequency (ITF), were conducted at electrode Cz, as the ErrP components prominently manifest over fronto-central areas like FCz and Cz [10].

To extract the ERN and Pe components, we first computed difference waveforms by subtracting the grand average of correct epochs from error epochs for each group (AIRTrode and research-grade gel) and session (morning offline and afternoon offline). Visual inspection of the grand average identified consistent timings for ERN and Pe across sessions and groups: ERN appeared as a negative deflection around 200 ms post-cursor onset, and Pe as a positive deflection around 260 ms post-onset, within a window from [150, 450] ms relative to the onset (figure 5). Therefore, three zero-crossing timestamps marking the start and end of these components were retained from this window.

On the grand average of each run, the Pe amplitude was defined as the maximum value observed between the first and last zero-crossing timestamps. Conversely, the ERN amplitude was determined as the minimum value between the first zero-crossing and the Pe peak. The peak-to-peak amplitude was calculated by subtracting the ERN amplitude from the Pe amplitude. Latencies for each component were measured from the time post-cursor onset to when the respective peaks occurred.

Since ErrPs are often associated with an increase in fronto-central θ power [10], we also computed the individual θ peak frequency (ITF) as a neurophysiological characteristic of ErrP. To compute the ITF, we first performed a time-frequency decomposition of single-trial EEG data using complex Morlet wavelets (1 to 30 Hz with 0.1 Hz resolution and 6 cycles for increased spectral resolution). After decomposition, we squared and averaged the magnitudes across trials to compute the average total power for each frequency at Cz, decibel normalized by the power during a baseline interval ([−500, −250] ms pre-cursor onset). The difference in average power between error and correct trials was averaged over a window of [200, 600] ms post-cursor onset, yielding the power per frequency. The ITF was identified as the frequency with the maximum power within the [4,8] Hz range. These neurophysiological characteristics–pk-to-pk amplitude, ERN latency, Pe latency, and ITF–were extracted from each participant's recording per run.

2.7. Statistical analyses

The BCI performance and neurophysiological results satisfied the assumptions of the statistical tests used. We assessed data normality using the Lilliefors test, considering data normally distributed if the p-value exceeded 0.05 [62]. We employed mixed-effects model to study the effects of time (morning vs. afternoon sessions) and electrode type (AIRTrode vs. research-grade gel) on BCI performance and neurophysiological patterns. For this analysis, we used each run as a separate data point, allowing us to utilize the full dataset and increase statistical power. The inclusion of a random effect for subjects models the correlation within subjects across multiple runs, validating the use of run-level data and preserving the model's statistical integrity. To evaluate the changes in BCI performances and neurophysiological patterns between morning and afternoon sessions, paired two-tailed t-tests were used for normally distributed data, and the Mann-Whitney U test was applied otherwise. Differences between electrode types were evaluated using unpaired two-tailed t-tests for normally distributed data and Wilcoxon signed-rank tests for non-normal data. Results were reported with means and standard deviations to describe central tendency and variability. Statistical significance was indicated with p-values, and effect sizes were calculated using Cohen's d_z alongside t-statistics (when applicable), providing insight into the magnitude of observed effects. Apart from the mixed-effect model analysis, all other statistical analyses of the BCI performances

and neurophysiological results were conducted at the subject level.

To explore the similarities in neurphysiological patterns, we performed Pearson's two-tailed correlation, and reported the correlation coefficients r and corresponding p-values. To adjust for multiple comparisons, we applied the Bonferroni correction by dividing the significance threshold by the number of tests in the corresponding analysis (i.e. $\alpha_{adjusted} = \frac{\alpha}{number\ of\ tests}$). Additionally, cluster permutation testing, implemented via FieldTrip, facilitated comparisons of spatial distributions of ERD patterns across electrode types and sessions, while also adjusting for multiple comparisons.

3. Results

3.1. An online, AIRTrode, MI-based BCI operates reliably for at least six hours

3.1.1. Neurophysiology

To evaluate the effectiveness of AIRTrodes in long-term MI-based BCIs, we first analyzed the ERD patterns over the contra-lateral C3/C4 channels and their topological distribution during the MI execution periods, in both μ and β bands.

As illustrated in figure 2, the grand average power patterns for both electrode types across all sessions exhibited a clear ERD (i.e. decrease in power) over the contra-lateral C3/C4 channels upon execution of the MI tasks, in the μ and β bands. Additionally, supplementary figure S2 showed physiologically valid lateralization of μ and β ERDs towards the contra-lateral hemisphere in the morning (morn) and afternoon (aft) online sessions across both electrode types. These observations aligned with established neurophysiological patterns of MI reported in the literature [45, 63].

Pearson's correlation analysis confirmed high similarities between the grand average ERD patterns at C3/C4 recorded from AIRTrodes and those from research-grade gel electrodes across both online sessions morn online: $r = 0.935$, $p < 0.001$ for μ, $r = 0.856$, $p < 0.001$ for β; aft online: $r = 0.899$, $p < 0.001$ for μ, $r = 0.884$, $p < 0.001$ for β), below adjusted α level of 0.0125 for each frequency band (i.e. 0.05 divided by 4 for Bonferroni correction). supplementary figure S2 confirmed that the topological distribution of ERD patterns showed mostly no significant differences across the two electrode types ($p > 0.05$). An exception was observed in supplementary figure S2b, where AIRTrode captured significantly stronger β ERDs over the contra-lateral C4 channel in the morning session when compared to gel ($t_{C4} = -4.959$, $p < 0.001$).

Furthermore, Pearson's correlation analysis showed strong correlations in the grand average ERD patterns at C3/C4 between the morning and afternoon sessions for AIRTrode (μ band: $r = 0.878$, $p < 0.001$; β band: $r = 0.953$, $p < 0.001$) consistent with the case for the standard gel (μ band: $r = 0.945$, $p < 0.001$; β band: $r = 0.918$, $p < 0.001$). The topological analysis through cluster permutation testing revealed no significant differences in ERD patterns between morning and afternoon sessions for AIRTrode, underscoring its stability ($p > 0.05$). Conversely, the research-grade gel electrodes showed a significant shift in μ ERD patterns on channel CP6 upon re-application in the afternoon session ($t_{CP6} = -3.195$, $p < 0.001$), suggesting potential variability with traditional electrodes.

Overall, these results support that AIRTrodes consistently captured physiologically valid, contra-lateral ERD patterns during MI-BCI operations over extended periods. These patterns were mostly not statistically different from those captured by research-grade gel electrodes, highlighting the reliability and stability of AIRTrode in long-term BCI operation. The only exception was when AIRTrode exhibited stronger contra-lateral SMR modulations than the standard gel during MI tasks at the morning session.

3.1.2. Online BCI performances

To assess the effectiveness of AIRTrode in supporting long-term MI-based BCI operations, figure 3 details the online classification accuracy, bar dynamics, and BCI hits accuracy of the MI-BCI across morning (morn) and afternoon (aft) sessions. For both electrode types, the performance metrics consistently exceeded their respective chance levels throughout the sessions.

A mixed-effect model with TIME (morn and aft) as a within-subjects factor and GROUP (AIRTrode and research-grade gel) as a between-subjects factor revealed no significant TIME x GROUP interaction effect (classification accuracy: $Coefficient = 2.744 \pm 4.929$, $t(44) = 0.557$, $p = 0.581$; bar dynamics: $Coefficient = 3.915 \pm 5.045$, $t(44) = 0.633$, $p = 0.530$, and BCI hits: $Coefficient = -5.417 \pm 9.157$, $t(44) = -0.592$, $p = 0.557$). In addition, there was no significant main effect of TIME (classification accuracy: $Coefficient = -3.025 \pm 3.485$, $t(44) = -0.868$, $p = 0.390$; bar dynamics: $Coefficient = -2.255 \pm 3.567$, $t(44) = -0.632$, $p = 0.531$; and BCI hits: $Coefficient = 2.083 \pm 6.475$, $t(44) = 0.322$, $p = 0.749$) nor GROUP (classification accuracy: $Coefficient = -0.970 \pm 4.828$, $t(44) = -0.201$, $p = 0.842$; bar dynamics: $Coefficient = -2.934 \pm 5.073$, $t(44) = -0.578$, $p = 0.566$; and BCI hits: $Coefficient = 3.333 \pm 13.778$, $t(44) = 0.242$, $p = 0.810$), indicating consistent performance across times and groups.

Post-hoc t-tests further confirmed no statistically significant differences in BCI performances between sessions for each electrode type. For AIRTrode, results showed no significant differences in classification accuracy ($morn = 62.058 \pm 8.927\%$,

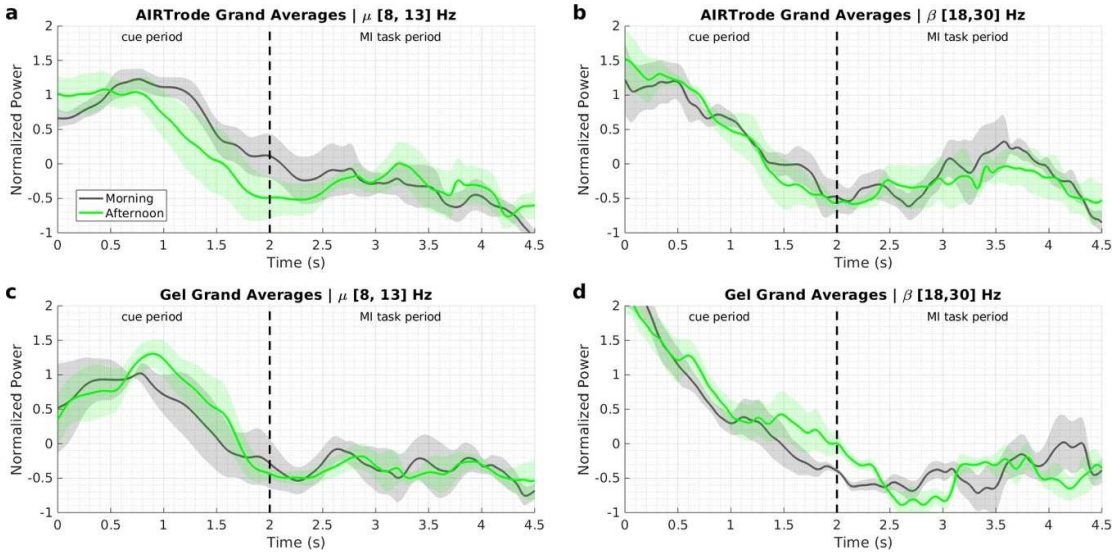

Figure 2. ERD grand averages in MI trials. Contra-lateral ERD patterns averaged over C3 and C4 channels for corresponding RH and LH trials in the morning and afternoon online sessions. *a* and *b* display the μ and β band ERDs, respectively, from AIRTrode recordings; *c* and *d* present the μ and β band ERDs from research-grade gel recordings. Shaded areas represent the standard error over participants [$n = 4$]. Horizontal black dashed lines indicate the start of the MI execution period following 2 s of cue presentation.

$aft = 65.084 \pm 8.997\%$, $p = 0.574$, $d_z = 0.338$), bar dynamics ($morn = 61.024 \pm 8.963\%$, $aft = 63.280 \pm 9.972\%$, $p = 0.875$, $d_z = 0.238$) and BCI hits accuracy ($morn = 60.417 \pm 15.237\%$, $aft = 62.727 \pm 39.402\%$, $p = 0.878$, $d_z = 0.084$). Similar non-significant variations across sessions were observed with the gel electrodes in classification accuracy ($morn = 63.832 \pm 10.848\%$, $aft = 64.114 \pm 5.076\%$, $t_3 = 0.046$, $p = 0.966$, $d_z = 0.033$), bar dynamics ($morn = 61.285 \pm 11.509\%$, $aft = 60.345 \pm 6.539\%$, $p = 0.896$, $d_z = -0.100$) and BCI hits accuracy ($morn = 58.333 \pm 11.982\%$, $aft = 61.667 \pm 8.001\%$, $p = 0.783$, $d_z = 0.164$).

Comparisons between electrode types in each session also showed no significant differences in the online classification accuracy (morn: $t_6 = -0.253$, $p = 0.809$, $d_z = 0.179$; aft: $t_6 = -0.188$, $p = 0.857$, $d_z = -0.133$), bar dynamics (morn: $t_6 = 0.036$, $p = 0.973$, $d_z = 0.024$; aft: $p = 0.486$, $d_z = 0.348$) and BCI hits accuracy (morn: $p = 0.893$, $d_z = -0.099$; aft: $p = 0.857$, $d_z = 0.133$). Additionally, analysis of the EEG channels contributing to BCI decoding performance confirmed significant contributions from similar, physiologically relevant, C3/C4 channels for both AIRTrode and research-grade gel, as depicted in figure 4. Subject-wise illustration of EEG channel contributions to MI decoding performance shows similar trends, despite some individual variations (supplementary figure S3).

Overall, the effect sizes (d_z) reported in the post-hoc tests for the BCI performance metrics between sessions and electrodes are classified as either trivial ($0.0 \leqslant d_z \leqslant 0.20$) or small ($0.20 \leqslant d_z \leqslant 0.50$), suggesting minor differences if any. These results, consistent with neurophysiological findings from section 3.1.1, confirm the robustness of AIRTrode for long-term BCI operations, maintaining reliable online BCI performances comparable to that of freshly applied gold-standard gel electrodes even after extended periods.

3.2. An online, AIRTrode, ErrP-based BCI operates reliably for at least six hours

3.2.1. Neurophysiology

To evaluate the robustness of AIRTrodes for long-term ErrP-driven BCIs, we first analyzed changes in the grand averages and specific neurophysiological characteristics of ErrPs at channel Cz across morning (morn) and afternoon (aft) sessions. The neurophysiological characteristics examined include: (1) peak-to-peak amplitude (pk-to-pk), computed as magnitude of Error Positivity (Pe) minus Error-related Negativity (ERN), (2) ERN latency, (3) Pe latency, and (4) the individual peak θ frequency of error trials (ITF).

The grand averages of the ErrPs from the error minus correct condition are displayed in figure 5. ErrPs recorded from both AIRTrode and gel electrodes in the morning and afternoon sessions displayed physiologically valid shapes in the temporal domain, aligning with established literature [11, 64, 65]. Consistently across both types of electrodes and both sessions, the topological plots revealed that the ErrPs were characterized by a central negative deflection appearing approx. 200 ms after the cursor onset of the erroneous action (ERN), followed by a frontal and fronto-central positive peak (Pe) approx. 260 ms after cursor onset. These patterns are established EEG neural correlates of error recognition [11, 64, 65]. An unexpected positive deflection at Fz, approx.

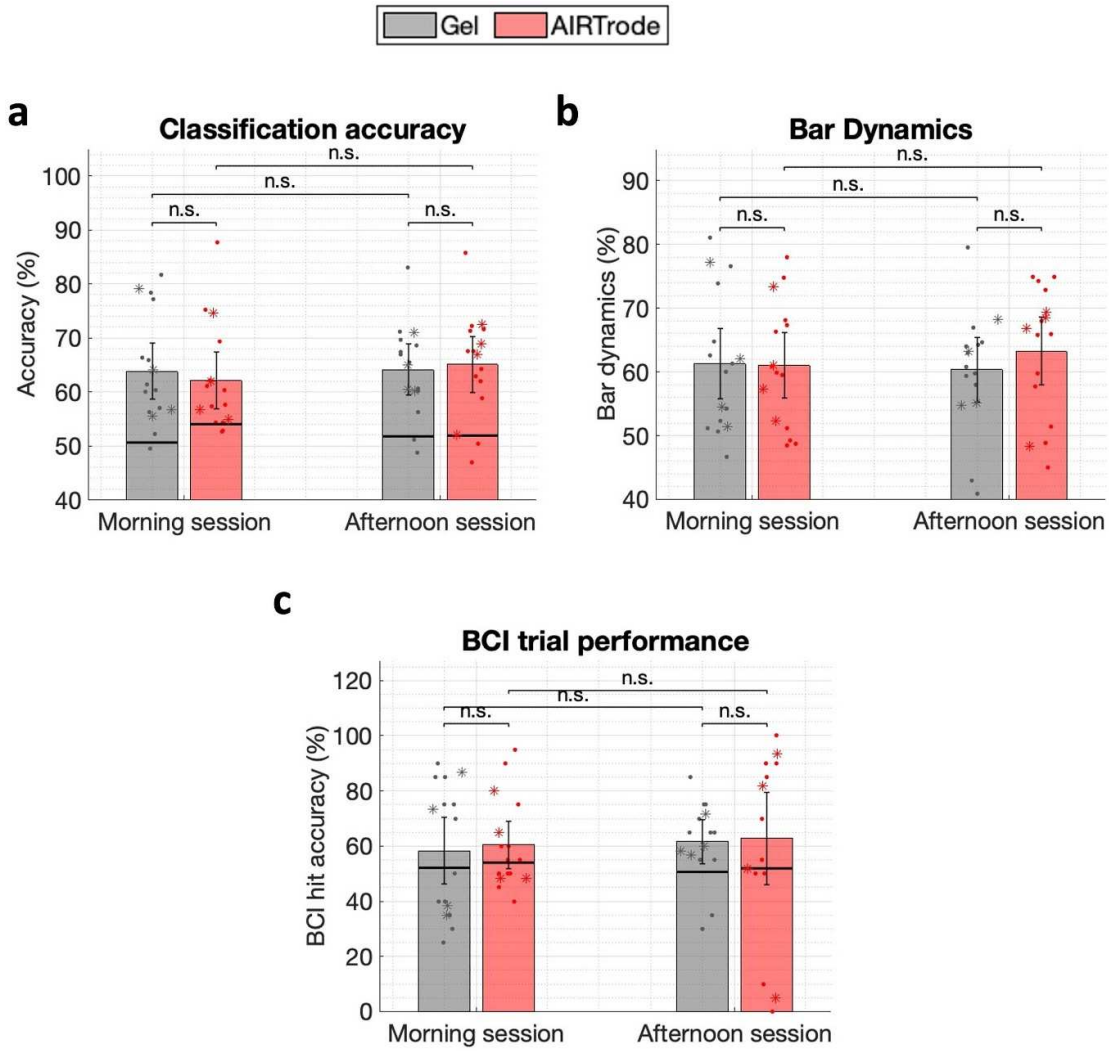

Figure 3. *Online MI-based BCI performance. a* Classification accuracy, *b* Bar dynamics, and *c* BCI hits accuracy in online MI-based BCI feedback for participants wearing AIRTrode [$n = 12$] and gel [$n = 12$] electrodes during morning and afternoon sessions. Each dot indicates a performance value from one run, with three runs per session. Asterisks indicate the average performance of each participant per session, calculated across three runs. There are four participants for each electrode type. Error bars represent the standard error of the mean for the performance metrics. Solid horizontal lines within the bars denote the chance levels for each metric and session. *n.s.* indicates not significant at the adjusted *alpha* level of 0.0125 (i.e. 0.05 divided by 4 for Bonferroni correction).

200 ms after error observation in the gel recordings, was likely due to ocular artifacts generated during the error monitoring task. Subject-wise illustration of ErrP grand averages show similar physiologically valid shapes, despite some individual variations (supplementary figure S4, S5).

Furthermore, strong Pearson's correlations were observed between the grand average ErrPs across sessions for both types of electrodes (AIRTrode, $r = 0.794$, $p < 0.001$; gel, $r = 0.747$, $p < 0.001$) and between the two electrodes within each session (morn offline, $r = 0.692$, $p < 0.001$; aft offline, $r = 0.821$, $p < 0.001$), below adjusted α level of 0.0125 (i.e. 0.05 divided by 4 for Bonferroni correction).

We conducted a mixed-effect analysis to evaluate each of the quantitative neurophysiological measures with TIME (morn and aft) as a within-subjects factor and GROUP (AIRTrode vs gel) as a between-subjects factor. Our analysis revealed no significant TIME X GROUP interaction effects in any physiological measures (pk-to-pk: $Coefficient = 0.176 \pm 0.419$, $t(44) = 0.421$, $p = 0.676$; ERN latency: $Coefficient = 0.007 \pm 0.023$, $t(44) = 0.291$, $p = 0.772$; Pe latency: $Coefficient = 0.001 \pm 0.018$, $t(44) = 0.082$, $p = 0.935$; ITF: $Coefficient = 0.233 \pm 0.726$, $t(44) = 0.322$, $p = 0.749$, indicating no group-specific changes across sessions. Additionally, no significant main effects of TIME were observed (pk-to-pk: $Coefficient = -0.202 \pm 0.96$, $t(44) = -0.683$, $p = 0.498$; ERN latency: $Coefficient = 0.001 \pm 0.016$, $t(44) = 0.050$, $p = 0.960$; Pe latency: $Coefficient = -0.003 \pm 0.012$, $t(44) = -0.245$, $p = 0.808$; ITF: $Coefficient = 0.333 \pm 0.513$, $t(44) = 0.650$, $p = 0.519$). Furthermore, no significant main effects of GROUP were observed (pk-to-pk: $Coefficient = -0.299 \pm 0.874$, $t(44) = -0.342$, $p = 0.734$; ERN

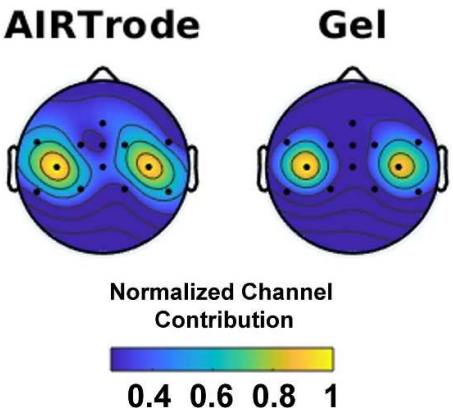

Figure 4. *EEG channel contributions to MI decoding performance.* Topological plots show the contribution of each recorded EEG channel to the performance of the Riemannian Geometry decoding for participants wearing AIRTrode and gel electrodes, averaged across all subjects and recorded sessions. The channel importance was computed as the percentage increase in classification accuracy when the channel was included in the classification model. Values were normalized relative to the maximum across all channels.

latency: $Coefficient = -0.001 \pm 0.016$, $t(44) = -0.050$, $p = 0.960$; Pe latency: $Coefficient = 0.001 \pm 0.018$, $t(44) = 0.051$, $p = 0.960$; ITF: $Coefficient = -0.367 \pm 0.513$, $t(44) = -0.715$, $p = 0.479$).

Post-hoc analyses further confirmed no significant changes across sessions for both electrode types in the physiological characteristics of ErrPs. For AIRTrode, results showed no significant differences in pk-to-pk amplitude ($morn = 2.415 \pm 1.497$ uV, $aft = 2.617 \pm 1.593$ uV, $t_3 = -0.320$, $p = 0.696$, $d_z = 0.131$), ERN latency ($morn = 0.184 \pm 0.013$ s, $aft = 0.183 \pm 0.029$ s, $p = 0.948$, $d_z = -0.037$), Pe latency ($morn = 0.263 \pm 0.017$ s, $aft = 0.266 \pm 0.034$ s, $p = 0.770$, $d_z = -0.115$) and ITF ($morn = 5.967 \pm 0.789$ Hz, $aft = 5.492 \pm 1.009$ Hz, $t_3 = 0.642$, $p = 0.567$, $d_z = -0.368$). Similarly, gel electrodes also exhibited no significant post-hoc differences across sessions, as demonstrated by pk-to-pk amplitude ($morn = 2.292 \pm 1.535$ uV, $aft = 2.318 \pm 1.101$ uV, $t_3 = -0.077$, $p = 0.944$, $d_z = 0.020$), ERN latency ($morn = 0.190 \pm 0.018$ s, $aft = 0.182 \pm 0.034$ s, $p = 0.449$, $d_z = -0.271$), Pe latency ($morn = 0.265 \pm 0.022$ s, $aft = 0.267 \pm 0.038$ s, $p = 1.0$, $d_z = 0.052$) and ITF ($morn = 5.542 \pm 0.578$ Hz, $aft = 5.342 \pm 0.691$ Hz, $t_3 = 1.897$, $p = 0.154$, $d_z = -0.890$).

In addition, no post-hoc differences were detected between the two electrode types in either session for any of the measured neurophysiological parameters: pk-to-pk (morn: $t_6 = 0.115$, $p = 0.912$, $d_z = -0.081$; aft: $t_6 = 0.309$, $p = 0.768$, $d_z = -0.218$), ERN latency (morn: $t_6 = -0.525$, $p = 0.618$, $d_z = 0.371$; aft: $p = 0.972$, $d_z = -0.026$), Pe latency (morn: $p = 0.914$, $d_z = 0.122$; aft: $p = 0.971$, $d_z = 0.027$), and ITF (morn: $t_6 = 0.273$, $p = 0.794$, $d_z = -0.193$; aft: $t_6 = 0.600$, $p = 0.571$, $d_z = -0.424$).

The effect sizes (d_z) reported in the post-hoc tests were mostly trivial ($0.0 \leqslant d_z \leqslant 0.20$) or small ($0.20 \leqslant d_z \leqslant 0.50$), suggesting any observed differences are minor and practically insignificant. Notably, a large effect size was observed in the within-group comparison of gel electrodes for ITF, which did not reach statistical significance. The large effect size observed in ITF within the gel group was primarily driven by two subjects who exhibited approx. a 1 Hz change in their peak theta frequency between the morning and afternoon sessions, while the other two subjects showed much smaller changes, remaining within 0.2 Hz. We hypothesise that this variability reflects typical neural fluctuations rather than a consistent effect of the electrode type. This is further supported by evidence in the literature showing that theta frequency dynamics can be influenced by a variety of factors, including internal thoughts, anxiety, and temperature [66, 67].

Together, these findings demonstrate the effectiveness of AIRTrodes in capturing event-related potentials, such as ErrPs. Moreover, the consistent absence of significant differences in ErrPs across time or between electrode types underscores AIRTrode's reliable and stable recording capabilities over extended periods.

3.2.2. Online BCI performances.

To evaluate the effectiveness of AIRTrode in supporting long-term ErrP-based BCI operations, figure 7 shows the online Kappa, TNR and TPR of the ErrP-BCI across morning (morn) and afternoon (aft) sessions. The performance metrics for both types of electrodes consistently exceeded their corresponding chance levels across sessions, apart from the TPR with the AIRTrode set-up was slightly below chance ($< 5\%$) in the afternoon session.

We performed a mixed-effect analysis on each of the BCI performance metrics with TIME (morn and aft) as within-subjects factors and GROUP (AIRTrode and gel) as between-subjects factor. Similar to the observations on the neurophysiological measures, no significant TIME x GROUP interaction effects were found in the performance measures (Kappa: $Coefficient = -0.012 \pm 0.067$, $t(44) = -0.184$, $p = 0.855$; TNR: $Coefficient = 0.035 \pm 0.070$, $t(44) = 0.500$, $p = 0.620$; TPR: $Coefficient = -0.020 \pm 0.035$, $t(44) = -0.573$, $p = 0.570$). Furthermore, there were no significant main effects of GROUP (Kappa: $Coefficient = -0.002 \pm 0.094$, $t(44) = -0.018$, $p = 0.986$; TNR: $Coefficient = -0.082 \pm 0.049$, $t(44) = -1.673$, $p = 0.101$; TPR: $Coefficient = 0.034 \pm 0.060$, $t(44) = 0.571$, $p = 0.571$.). There was also no significant main effects of TIME (TNR: $Coefficient = 0.037 \pm 0.049$, $t(44) = 0.759$, $p = 0.452$; TPR: $Coefficient = 0.031 \pm 0.024$, $t(44) =$

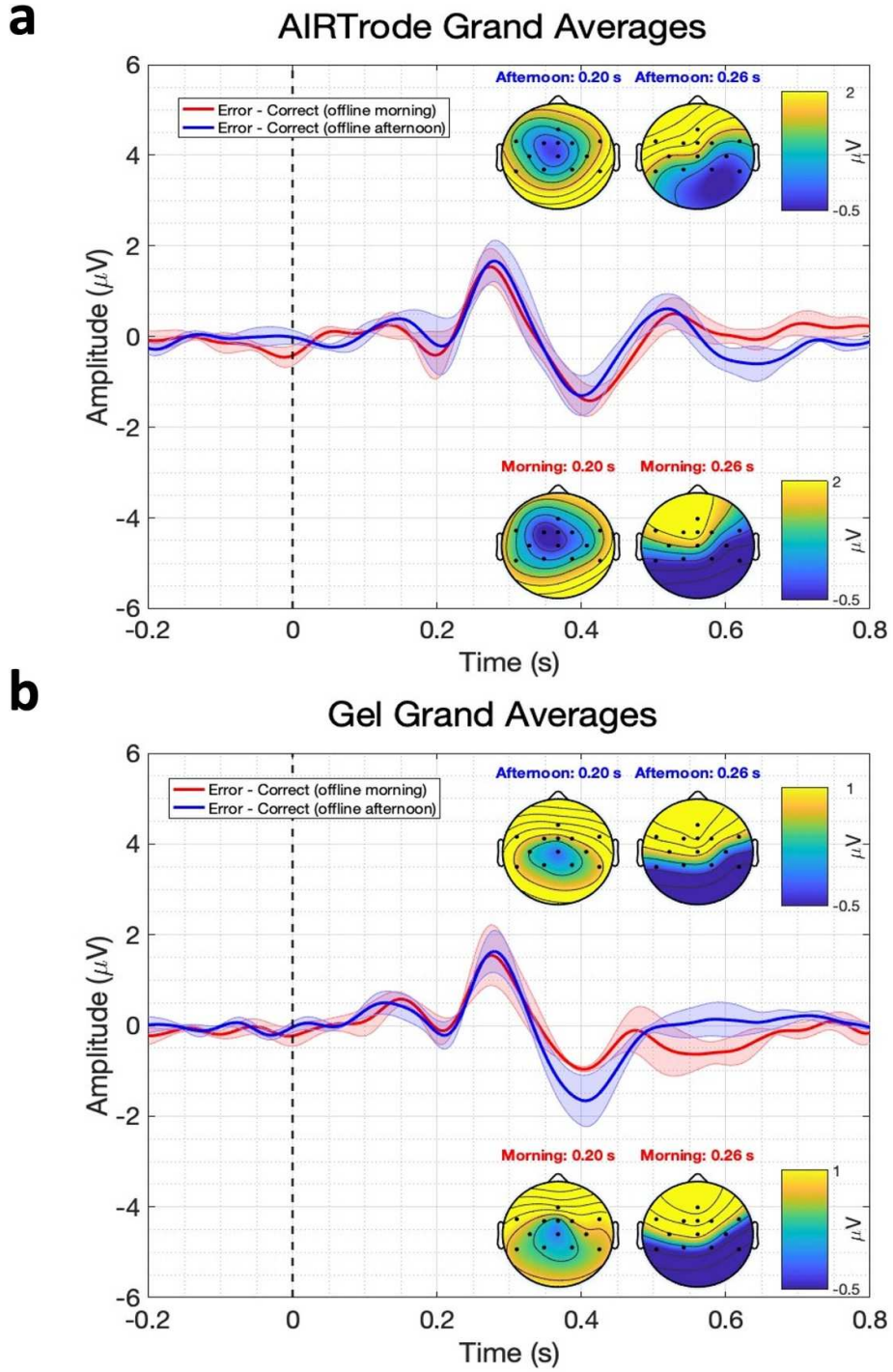

Figure 5. *ErrP grand averages. a* Grand averages of time-locked ErrPs, error-minus-correct condition, at Cz recorded with hydrogel AIRTrodes during morning and afternoon offline sessions. Time 0 s, marked by horizontal black dashed line, corresponds to the onset of cursor movement. Shaded areas represent the standard error over participants [$n = 4$]. Inset is topographical plot of EEG amplitude over the participants' scalp for ErrPs at two different time points with respect to the cursor onset: 0.20 s and 0.26 s. ***b*** Response-locked grand averages of ErrPs as described in *a*, recorded using research-grade gel electrodes.

1.267, $p = 0.212$; Kappa: *Coefficient* $= 0.073 \pm 0.047$, $t(44) = 1.542$, $p = 0.130$).

Post-hoc t-tests confirmed that changes in BCI performance across sessions were not significant for either electrode type. For AIRTrode, results showed no significant differences in Kappa ($morn = 0.390 \pm 0.095\%$, $aft = 0.317 \pm 0.160\%$, $t_3 = 1.458$, $p = 0.241$, $d_z = -0.558$), TNR ($morn = 71.209 \pm 9.270\%$, $aft = 67.480 \pm 9.454\%$, $p = 0.666$, $d_z = -0.398$) and TPR ($morn = 72.627 \pm 6.214\%$, $aft = 69.526 \pm 8.9.018\%$, $t_3 = 1.086$, $p = 0.357$, $d_z = -0.400$). For gel, no significant differences were observed in Kappa ($morn = 0.375 \pm 0.156\%$, $aft = 0.315 \pm 0.173\%$, $t_3 = 2.541$, $p = 0.085$, $d_z = -0.369$), TNR ($morn = 66.462 \pm 7.720\%$, $aft = 59.258 \pm 4.681\%$, $p = 0.106$, $d_z = -1.128$) and

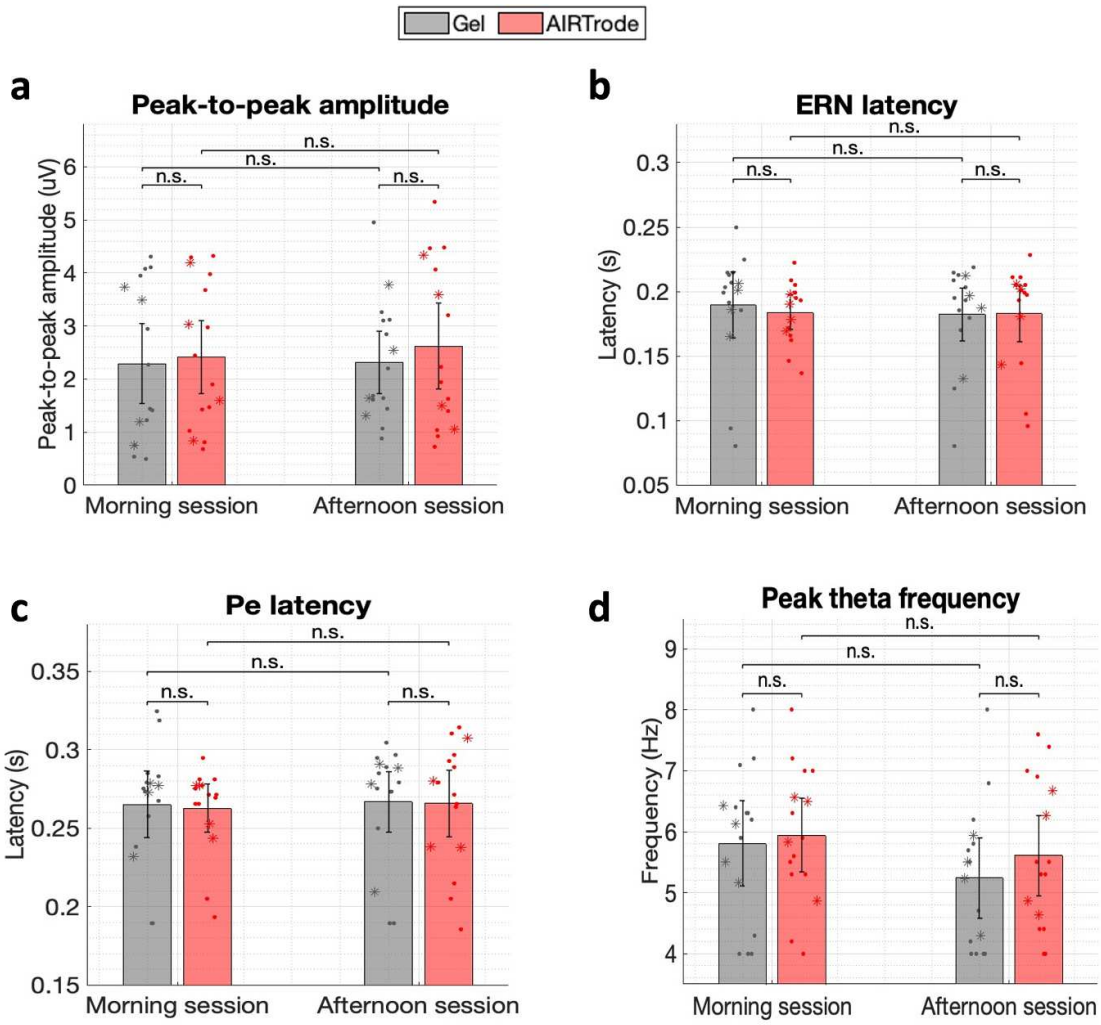

Figure 6. *Neurophysiological properties of ErrPs. a* Peak-to-peak amplitude, *b* ERN latency, *c* Pe latency, and *d* peak θ frequency of the ErrPs recorded during the offline morning and afternoon sessions from participants wearing AIRTrode [$n = 12$] and gel [$n = 12$] electrodes. Each dot represents a neurophysiological value from one of three runs per session, Asterisks indicate the average performance of each participant per session, calculated across three runs. There are four participants for each electrode type. Error bars represent the standard error of the mean for the performance metrics. *n.s.* indicates not significant at the adjusted *alpha* level of 0.0125 (i.e. 0.05 divided by 4 for Bonferroni correction).

TPR ($morn = 74.066 \pm 10.690\%$, $aft = 72.947 \pm 13.500\%$, $p = 0.875$, $d_z = -0.093$).

Additionally, no significant differences were found between the two groups at any of the metrics or sessions in the independent samples post-hoc t-tests: Kappa (morn: $t_6 = 0.153$, $p = 0.883$, $d_z = -0.108$; aft: $t_6 = 0.014$, $p = 0.989$, $d_z = -0.009$), TNR (morn: $p = 0.461$, $d_z = -0.556$; aft: $p = 0.170$, $d_z = -1.102$), and TPR (morn: $t_6 = -0.244$, $p = 0.816, d_z = 0.172$; aft: $p = 0.686$, $d_z = 0.295$).

When evaluating BCI performance in this paradigm, we recommend prioritizing the Kappa value, which accounts for both TNR and TPR, providing a more balanced assessment. For comprehensiveness, we reported the statistical results for all three metrics above. Cross-group comparisons between electrode types showed that AIRTrode consistently outperformed gel electrodes in the TNR metric, with a medium effect size observed in the morning session and a strong effect size in the afternoon. However, this comparison should be interpreted in the context of TPR as well, where AIRTrode consistently showed a lower average performance. Therefore, considering the Kappa value, which integrates both TNR and TPR, offers a more balanced evaluation of BCI performance. The effect sizes (d_z) reported in the post-hoc tests of Kappa are classified as either trivial ($0.0 \leqslant d_z \leqslant 0.20$) or small ($0.20 \leqslant d_z \leqslant 0.50$), apart from the Kappa value for AIRTrode showed a non-significant decline with a nearly medium effect size across sessions. It remained high and well above chance levels despite the slight decrease.

In summary, the performance of an ErrP-based BCI employing AIRTrode electrodes was generally reliable and comparable to that of research-grade gel electrodes, although the TPR slightly dipped below chance levels in the afternoon session, possibly due to participant fatigue. Nonetheless, this value remained high because the task design, where only

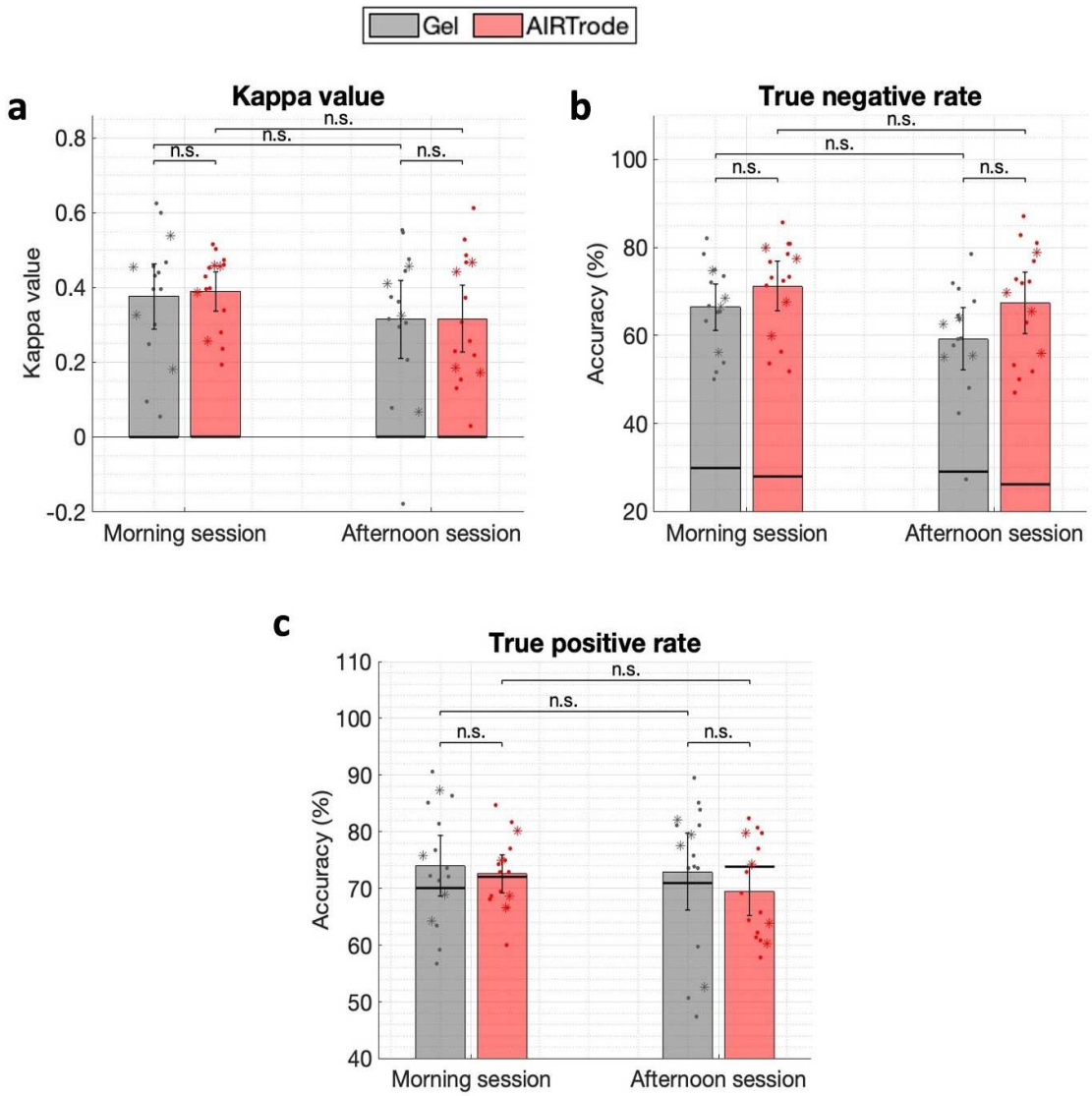

Figure 7. *Online ErrP-based BCI performance. a* Kappa value, *b* True negative rate (TNR), and *c* True positive rate (TPR) in online ErrP-based feedback for participants wearing AIRTrode [$n = 12$] and gel [$n = 12$] electrodes during morning and afternoon sessions. Each dot indicates a performance value from one run, with three runs per session. Asterisks indicate the average performance of each participant per session, calculated across three runs. There are four participants for each electrode type. The correct trials were labelled positive and the error trials were labelled negative. Error bars represent the standard error of the mean for the performance metrics. Solid horizontal lines within the bars denote the chance levels for each metric and session. *n.s.* indicates not significant at the adjusted *alpha* level of 0.0125 (i.e. 0.05 divided by 4 for Bonferroni correction).

30% of cursor movements were erroneous to create an expectation mismatch, sets a high chance level for TPR ($> 70\%$). Overall, these findings confirm the effectiveness and stability of AIRTrode for long-term use in ErrP-based BCIs.

4. Discussion

In this work, we have demonstrated three key aspects of hydrogel AIRTrode in the context of EEG BCI operations, underscoring its potential promise for future clinical and practical applications. Firstly, the AIRTrode electrodes exhibited excellent recording quality, effectively capturing valid neurophysiological patterns that characterize MI and ErrPs. For MI tasks, the electrodes consistently tracked continuous oscillatory SMRs, while for ErrPs, they reliably captured discrete, event-related potentials that occur in response to specific errors. In addition, the impedance of AIRTrodes was comparable to that of research-grade gel electrodes, with no significant differences in average impedance between the two (morn: $t_6 = 1.160$, $p = 0.290$, $d_z = 0.820$; aft: $t_6 = 1.517$, $p = 0.180$, $d_z = 1.072$)(supplementary figure S6). Though the average impedance across subjects was slightly higher in AIRTrode than that of gel electrodes with a large effect size, it remained below 60 kΩ for all participants and sessions, indicating good recording quality.

Secondly, AIRTrodes maintained stability in recording quality over extended periods–at least

six hours. There were no significant differences in average impedance over the two sessions, and medium effect size (morn: 29.800 ± 3.747 kΩ, aft: 36.133 ± 12.569 kΩ, $t_3 = 1.263, p = 0.296$, $d_z = 0.683$). This medium effect size was primarily driven by a single subject whose impedance increased by 20 kΩ, while the impedance variation for the remaining subjects stayed within 5.5 kΩ. This stability was also demonstrated in that the characteristics of the ErrPs (figures 5 and 6(a)–(d)) and the ERD patterns of MI (figure 2 and supplementary figure S2) showed no significant changes over the extended period. This stability minimizes the need for frequent re-calibrations, which can be cumbersome and time-consuming in everyday use.

Thirdly and most importantly, the combination of high-quality recording and prolonged stability underpins the reliable functionality of AIRTrode in online BCI operation for extended durations. For both tasks, online performances in both sessions were reliable and comparable to that achieved with research-grade gel electrodes. While our findings suggest that gel-based systems can maintain satisfactory BCI performances even after re-application, the frequent need for re-application poses practical challenges. It can be time-consuming, uncomfortable, and inconvenient for users in real-world settings.Comparatively to gel electrodes, AIRTrode's reduced need for consumables and less frequent maintenance/re-applications might potentially lead to lower operational costs including the costs of labor, which might often be the most significant expense in clinical and practical settings [68]. Its maintenance-free nature in continuous recordings may be particularly advantageous in dynamic environments, simplifying logistics and reducing activity interruptions. Most importantly, AIRTrode's stable, non-displacement design continuously maintains signal integrity, mitigating risks associated with frequent electrode adjustments. This study represents the first demonstration of a hydrogel electrode's effectiveness in supporting multi-hour BCI operations, setting a precedent for its application in long-term scenarios.

The direct implications of our research are significant. AIRTrode electrodes may be integrated into MI-based BCI-controlled devices, such as wheelchairs and robotic prostheses, and ErrP-based BCI spellers. These applications, as referenced in prior studies [5, 6, 11], may benefit in terms of practicality from our system's ability to operate reliably over extended periods, far surpassing typical durations in controlled laboratory settings. Furthermore, AIRTrodes hold potential promise for improving the effectiveness of BCI-based neuro-rehabilitation. The electrodes' continuous stability can help BCI users to learn to effectively modulate their SMRs across multiple longitudinal training sessions [69–71]. This sustained interaction can promote increased motor cortex engagement, enhanced corticospinal excitability [72–76], stronger muscle activations [73, 77, 78] and improved BCI performances [70], thereby supporting superior rehabilitation outcomes. Lastly, these electrodes show potential in BCI-controlled robotic assistants, as highlighted by the systems or concepts described in [79] and [80], which can benefit from extended operational periods.

In addition to BCI applications, the demonstrated stability and reliability of AIRTrodes may support their use in long-term, multi-hour EEG studies, such as sleep monitoring and the diagnosis of sleep disorders [34, 81]. Furthermore, these electrodes may be potentially suited for multi-day EEG longitudinal studies that track the progression of cognitive decline associated with healthy aging [82], as well as the progression of neurodegenerative diseases and their associated biomarker changes. Such studies include the monitoring of specific conditions like Alzheimer's disease [83, 84], multiple sclerosis [85], and ataxias [86]. Additionally, this capability holds potential promise for not only the long-term evaluation of therapeutic interventions, such as in stroke recovery [87], but also aids in the early detection of patients at risk of neurodegeneration [84]. Furthermore, AIRTrode's capability for long-term, high-quality EEG recordings may hold promise in enhancing personalized medicine by enabling continuous monitoring and treatment customization based on individual patient responses, particularly in complex cases like stroke recovery, as highlighted by Fleury *et al* [88].

Comparative analysis of BCI performances between AIRTrode and gel electrodes across two sessions, separated by a 6-hour interval, reveals no consistent performance advantage for either electrode type. In the MI task, while there are slight indications that the AIRTrode may outperform the gel electrode in session 2, these differences are not statistically significant ($P > 0.40$ with a small effect size of $d_z < 0.3$) across all performance metrics. Similarly, in the ErrP task, AIRTrode shows a slight edge over the gel electrode in session 1 when considering the Kappa metric, which accounts for both TNR and TPR, but again, this difference does not reach significance ($P > 0.40$ with a small effect size of $d_z < 0.3$). These preliminary observations suggest a potential trend favoring AIRTrode in online BCI performances; however, to confirm consistent performance advantages, these trends would need validation in studies with larger and more diverse populations.

Despite the novelties, the study has several limitations that should be acknowledged. Due to the demanding nature of the study and the commitment required, our sample size was small, and coincidentally, only male participants were enrolled despite efforts to recruit female participants. We recognize that this limitation may affect the generalizability of our findings to a broader population, including females. Future research should aim to

include a larger and more diverse sample, incorporating both male and female participants, to enhance the generalizability of the results. However, we do not anticipate significant differences in our results with female subjects for several reasons. Firstly, the hydrogel electrodes used in this study have demonstrated stable performance across both male and female subjects with various hair characteristics in pilot recordings [34]. Secondly, a recent comprehensive study involving over 200 subjects found no significant differences in neurophysiological characteristics during operation of MI-based BCIs, similar to the MI-BCI studied here, suggesting that sex may not be a predictive factor for MI-BCI performance [89]. While a prior study has indicated potential small differences in the amplitude of error-related brain activities between sexes [90], there is no evidence of sex-related discrepancies in the operation of ErrP-based BCIs. Furthermore, our male participants exhibited a diverse range of hair lengths, textures, and thicknesses, resembling those commonly found in the female population. This diversity included long curly hair, dense medium-length hair and fine medium-length hair, providing a valuable range of conditions for assessing the performance of the novel electrodes. While we do not expect significant differences in outcomes with female subjects, further studies are needed to confirm this with a balanced participant pool. Furthermore, despite the findings from this study, future research is needed to confirm its clinical and practical relevance.Additionally, while our findings indicate that AIRTrode may offer stability and user convenience, a comprehensive economic analysis is needed to fully assess its cost-effectiveness in practical and clinical settings. Future studies should compare the costs of AIRTrode with other electrodes, considering manufacturing and maintenance expenses, to evaluate its economic feasibility and potential for broader adoption. This comparative analysis should also consider parameters like signal quality to better assess AIRTrode's overall performance. To support this, we have included a comparison table in supplementary figure S7, which evaluates AIRTrode against other novel electrode types based on MI-BCI performance, ErrP-BCI performance, signal quality, long-term recording capability and user comfort/skin irritation. Lastly, future studies should validate the use of AIRTrode in BCI applications that extend beyond 6 hours, including operations spanning multiple days.

In conclusion, our findings demonstrate that hydrogel AIRTrode electrodes provide a reliable and stable interface for online EEG-based BCI operations, suitable for both short-term and extended use. This capability may mark a significant advancement toward practical, all-day BCI solutions that can be seamlessly integrated into everyday activities, offering sustained assistance and enhancing quality of life for users.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary figures).

Acknowledgments

H W would like to acknowledge support from UT Austin Proof of Concept Award. J H would like to acknowledge support from the Texas Innovation Center at the University of Texas at Austin. The authors also extend their gratitude to the research participants involved in this study. Additionally, the authors confirm that there are no conflicts of interest concerning the publication of this paper. [14, 23, 24, 30, 31, 33, 34, 91–99].

Author contributions statement

D L, J H, H A, H W, and J d R M conceived and designed the experimental protocols. D L, J H, H A, S K, F I, I P, and Z A implemented the experimental protocol. D L, J H, and H A conducted the experiments. D L, J H, and H A analysed the results. D L, J H, and H A prepared the draft manuscript. All authors reviewed and approved the final version of the manuscript.

ORCID iDs

Deland H Liu https://orcid.org/0000-0002-9293-4737
Ju-Chun Hsieh https://orcid.org/0000-0001-5598-5917
Fumiaki Iwane https://orcid.org/0000-0002-9659-4127
Ilya Pyatnitskiy https://orcid.org/0000-0002-2827-1473
José del R Millán https://orcid.org/0000-0001-5819-1522

References

[1] Wolpaw J R, Millán J del R and Ramsey N F 2020 Brain-computer interfaces: Definitions and principles *Brain-Computer Interfaces. Handbook of Clinical Neurology* vol 168, ed J del R Millán and N F Ramsey (Elsevier) pp 15–23
[2] Carpi F, De Rossi D and Menon C 2006 Non invasive brain-machine interfaces *ESA Ariadna Study* **5** 6402
[3] Tonin L and Millán J del R 2021 Noninvasive brain-machine interfaces for robotic devices *Annu. Rev. Control, Robot. Autonomous Sys.* **4** 191–214
[4] Decety J 1996 The neurophysiological basis of motor imagery *Behavioural Brain Res.* **77** 45–52
[5] Tonin L, Perdikis S, Deniz Kuzu T, Pardo J, Orset B, Lee K, Aach M, Armin Schildhauer T, Martínez-Olivera R and Millán J del R 2022 Learning to control a BMI-driven wheelchair for people with severe tetraplegia *iScience* **25** 105418
[6] Rupp R, Rohm M, Schneiders M, Kreilinger A and Müller-Putz G R 2015 Functional rehabilitation of the paralyzed upper extremity after spinal cord injury by

noninvasive hybrid neuroprostheses *Proc. of the IEEE* vol 103 pp 954–68
[7] Lee K, Liu D, Perroud L, Chavarriaga R and Millán J del R 2017 A brain-controlled exoskeleton with cascaded event-related desynchronization classifiers *Robot. Auton. Syst.* **90** 15–23
[8] Leeb R, Tonin L, Rohm M, Desideri L, Carlson T and Millán J d R 2015 Towards independence: a BCI telepresence robot for people with severe motor disabilities *Proc. of the IEEE* vol 103 pp 969–82
[9] Williamson J, Murray-Smith R, Blankertz B, Krauledat M and Millán J d R 2009 Designing for uncertain, asymmetric control: interaction design for brain-computer interfaces *Int. J. Hum. Comput. Studies* **67** 827–41
[10] Chavarriaga R, Sobolewski A and Millán J D R 2014 Errare machinale EST: the use of error-related potentials in brain-machine interfaces *Frontiers Neurosci.* **8** 208
[11] Chavarriaga R, Iturrate I naki and Millán J del R 2016 Robust, accurate spelling based on error-related potentials *6th Int. Brain-Computer Interface Meeting*
[12] Batzianoulis I, Iwane F, Wei S, Gaspar Pinto Ramos Correia C, Chavarriaga R, Millán J del R and Billard A 2021 Customizing skills for assistive robotic manipulators, an inverse reinforcement learning approach with error-related potentials *Commun. Biol.* **4** 1406
[13] Portillo-Lara R, Tahirbegi B, Chapman C A R, Goding J A and Green R A 2021 Mind the gap: state-of-the-art technologies and applications for EEG-based brain-computer interfaces *APL Bioeng.* **5** 031507
[14] Hua H, Tang W, Xiangmin X, Dagan Feng D and Shu L 2019 Flexible multi-layer semi-dry electrode for scalp EEG measurements at hairy sites *Micromachines* **10** 518
[15] Hsieh J-C, Yang Li, Wang H, Perz M, Tang Q, Kevin Tang K W, Pyatnitskiy I, Reyes R, Ding H and Wang H 2022 Design of hydrogel-based wearable EEG electrodes for medical applications *J. Mater. Chem.* B **10** 7260–80
[16] Angel Lopez-Gordo M, Sanchez-Morillo D and Pelayo Valle F 2014 Dry EEG electrodes *Sensors* **14** 12847–70
[17] Yang L, Liu Q, Zhang Z, Gan L, Zhang Y and Jinglong W 2022 Materials for dry electrodes for the electroencephalography: advances, challenges, perspectives *Adv. Mater. Technol.* **7** 2100612
[18] Harati A and Jahanshahi A 2021 A reliable stretchable dry electrode for monitoring of EEG signals *Sens. Actuators A: Phys.* **326** 112727
[19] Xing X, Wang Y, Pei W, Guo X, Liu Z, Wang F, Ming G, Zhao H, Gui Q and Chen H 2018 A high-speed SSVEP-based BCI using dry EEG electrodes *Sci. Rep.* **8** 14708
[20] Salvo P, Raedt R, Carrette E, Schaubroeck D, Vanfleteren J and Cardon L 2012 A 3D printed dry electrode for ECG/EEG recording *Sens. Actuators A: Phys.* **174** 96–102
[21] Mahmood M *et al* 2021 Wireless soft scalp electronics and virtual reality system for motor imagery-based brain-machine interfaces *Adv. Sci.* **8** 2101129
[22] Junshi Li *et al* 2022 High-performance flexible microneedle array as a low-impedance surface biopotential dry electrode for wearable electrophysiological recording and polysomnography *Nano-Micro Lett.* **14** 132
[23] Xue H *et al* 2023 Hydrogel electrodes with conductive and substrate-adhesive layers for noninvasive long-term EEG acquisition *Microsyst. Nanoeng.* **9** 79
[24] Min Lee S M, Hun Kim J H, Park C, Hwang J-Y, Sook Hong J S, Ho Lee K H and Hoon Lee S H 2015 Self-adhesive and capacitive carbon nanotube-based electrode to record electroencephalograph signals from the hairy scalp *IEEE Trans. Biomed. Eng.* **63** 138–47
[25] Garg R *et al* 2023 Wearable High-Density MXene-Bioelectronics for Neuromuscular Diagnostics, Rehabilitation and Assistive Technologies *Small Methods* **7** 2201318
[26] Driscoll N *et al* 2021 Mxene-infused bioelectronic interfaces for multiscale electrophysiology and stimulation *Sci. Trans. Med.* **13** eabf8629
[27] Driscoll N, Dong R and Vitale F 2021 Emerging approaches for sensing and modulating neural activity enabled by nanocarbons and carbides *Curr. Opin. Biotechnol.* **72** 76–85
[28] Hsieh J-C *et al* 2022 A highly stable electrode with low electrode-skin impedance for wearable brain-computer interface *Biosens. Bioelectron.* **218** 114756
[29] Guangli G, Wang S, Li M and Duan Y Y 2021 Towards real-life EEG applications: Novel superporous hydrogel-based semi-dry EEG electrodes enabling automatically charge-discharge' electrolyte *J. Neural. Eng.* **18** 046016
[30] Sheng X, Qin Z, HaiPeng X, Shu X, GuoYing G and Zhu X 2021 Soft ionic-hydrogel electrodes for electroencephalography signal recording *Sci. China Technol. Sci.* **64** 273–82
[31] Shen G, Gao K, Zhao N, Zhiran Y, Jiang C, Yang B and Liu J 2021 A novel flexible hydrogel electrode with a strong moisturizing ability for long-term EEG recording *J. Neural Eng.* **18** 066047
[32] Liu J, Lin S, Wenzheng Li, Zhao Y, Liu D, Zhaofeng H, Wang D, Lei M, Hong B and Hui W 2022 Ten-hour stable noninvasive brain-computer interface realized by semidry hydrogel-based electrodes *Research* **2022** 9830457
[33] Wang C *et al* 2022 On-skin paintable biogel for long-term high-fidelity electroencephalogram recording *Sci. Adv.* **8** eabo1396
[34] Hsieh J-C *et al* 2024 Design of an injectable, self-adhesive and highly stable hydrogel electrode for sleep recording *Device* **2** 100182
[35] Chen Y-H, Op De Beeck M, Vanderheyden L, Carrette E, Mihajlović V, Vanstreels K, Grundlehner B, Gadeyne S, Boon P and Van Hoof C 2014 Soft, comfortable polymer dry electrodes for high quality ecg and eeg recording *Sensors* **14** 23758–80
[36] Tianyou Y, Xiao J, Wang F, Zhang R, Zhenghui G, Cichocki A and Yuanqing Li 2015 Enhanced motor imagery training using a hybrid BCI with feedback *IEEE Trans. Biomed. Eng.* **62** 1706–17
[37] Alimardani M, Nishio S and Ishiguro H 2018 Brain-computer interface and motor imagery training: The role of visual feedback and embodiment *Evolving BCI Ther. Engaging Brain State Dyn.* **2** 64
[38] Neuper C, Scherer R, Wriessnegger S and Pfurtscheller G 2009 Motor imagery and action observation: modulation of sensorimotor brain rhythms during mental control of a brain–computer interface *Clin. Neurophysiol.* **120** 239–47
[39] Yang Chin Z, Keng Ang K, Wang C and Guan C 2010 Online performance evaluation of motor imagery BCI with augmented-reality virtual hand feedback *2010 Annual Int. Conf. of the IEEE Engineering in Medicine and Biology* (IEEE) pp 3341–4
[40] Leeb R, Perdikis S, Tonin L, Biasiucci A, Tavella M, Creatura M, Molina A, Al-Khodairy A, Carlson T and Millán J del R 2013 Transferring brain-computer interfaces beyond the laboratory: successful application control for motor-disabled users *Artif. Intell. Med.* **59** 121–32
[41] Alawieh H, Liu D, Madera J, Kumar S, Samuel Racz F, Majewicz Fey A and Millán J del R 2024 Transcutaneous electrical spinal cord stimulation promotes focal sensorimotor activation that accelerates brain-computer interface skill learning *medRxiv* 2024–06
[42] Samuel Racz F, Kumar S, Kaposzta Z, Alawieh H, Hu Liu D, Liu R, Czoch A, Mukli P and Millán J del R 2024 Combining detrended cross-correlation analysis with riemannian geometry-based classification for improved brain-computer interface performance *Frontiers Neurosci.* **18** 1271831
[43] McFarland D J 2000 Mu and beta rhythm topographies during motor imagery and actual movements *Brain topography* **12** 177–86

[44] Pfurtscheller G and Lopes Da Silva F H 1999 Event-related EEG/MEG synchronization and desynchronization: basic principles *Clin. Neurophysiol.* **110** 1842–57
[45] Pfurtscheller G 2000 Spatiotemporal ERD/ERS patterns during voluntary movement and motor imagery *Suppl. Clin. Neurophysiology* **53** 196–8
[46] Pfurtscheller G and Neuper C 2001 Motor imagery and direct brain-computer communication *Proc. of the IEEE* vol 89 pp 1123–34
[47] Barachant A, Bonnet S, Congedo M and Jutten C 2010 Riemannian geometry applied to BCI classification *Latent Variable Analysis and Signal Separation*
[48] Barachant A, Bonnet S, Congedo M and Jutten C 2011 Multiclass brain–computer interface classification by Riemannian geometry *IEEE Trans. Biomed. Eng.* **59** 920–8
[49] Congedo M, Barachant A and Bhatia R 2017 Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review *Brain Comput. Int.* **4** 155–74
[50] Kumar S, Alawieh H, Samuel Racz F, Fakhreddine R and Millán J del R 2024 Transfer learning promotes acquisition of individual BCI skills *PNAS Nexus* **3** gae076
[51] Zanini P, Congedo M, Jutten C, Said S and Berthoumieu Y 2017 Transfer learning: a riemannian geometry framework with applications to brain–computer interfaces *IEEE Trans. Biomed. Eng.* **65** 1107–16
[52] Pan L, Wang K, Lichao X, Sun X, Weibo Y, Minpeng X and Ming D 2023 Riemannian geometric and ensemble learning for decoding cross-session motor imagery electroencephalography signals *J. Neural Eng.* **20** 066011
[53] Ehrlich S K and Cheng G 2019 A feasibility study for validating robot actions using EEG-based error-related potentials *Int. J. Soc. Robot.* **11** 271–83
[54] Iturrate I, Montesano L and Minguez J 2013 Shared-control brain-computer interface for a two dimensional reaching task using EEG error-related potentials *2013 35th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE) pp 5258–62
[55] Kumar A, Pirogova E and Fang J Q 2018 Classification of error-related potentials using linear discriminant analysis *2018 IEEE-EMBS Conf. on Biomedical Engineering and Sciences (IECBES)* (IEEE) pp 18–21
[56] Ferrez P W and Millán J del R 2008 Simultaneous real-time detection of motor imagery and error-related potentials for improved BCI accuracy *Proc. of the 4th Int. Brain-Computer Interface Workshop and Training Course* pp 197–202
[57] Iturrate I naki, Grizou J, Omedes J, Oudeyer P-Y, Lopes M and Montesano L 2015 Exploiting task constraints for self-calibrated brain-machine interface control using error-related potentials *PLoS One* **10** e0131491
[58] Spüler M, Walter A, Rosenstiel W and Bogdan M 2014 Spatial filtering based on canonical correlation analysis for classification of evoked or event-related potentials in EEG data *IEEE Trans. Neural Sys. Rehabil. Eng.* **22** 1097–103
[59] Welch P 1967 The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms *IEEE Trans. Audio Electroacoust.* **15** 70–73
[60] Lopes-Dias C, Sburlea A I and Müller-Putz G R 2019 Online asynchronous decoding of error-related potentials during the continuous control of a robot *Sci. Rep.* **9** 17596
[61] Cohen J 1968 Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit *Psychol. Bull.* **70** 213
[62] Mendes M M and Pala A P 2003 Type i error rate and power of three normality tests *Pak. J. Inf. Technol.* **2** 135–9
[63] Jeon Y, Nam C S, Kim Y-J and Cheol Whang M 2011 Event-related (de) synchronization (ERD/ERS) during motor imagery tasks: Implications for brain-computer interfaces *Int. J. Ind. Ergon.* **41** 428–36
[64] Chavarriaga R and Millán J del R 2010 Learning from EEG error-related potentials in noninvasive brain-computer interfaces *IEEE Trans. Neural Sys. Rehabil. Eng.* **18** 381–8
[65] Ferrez P W and Millan J del R 2008 Error-related EEG potentials generated during simulated brain-computer interaction *IEEE Trans. Biomed. Eng.* **55** 923–9
[66] Hua J, Wolff A, Zhang J, Yao L, Zang Y and Luo J 2022 Xianliang Ge, Chang Liu and Georg Northoff. Alpha and theta peak frequency track on-and off-thoughts *Commun. Biol.* **5** 209
[67] Hines M, Poulter S, Douchamps V, Pibiri F, McGregor A and Lever C 2023 Frequency matters: how changes in hippocampal theta frequency can influence temporal coding, anxiety-reduction and memory *Frontiers Sys. Neurosci.* **16** 998116
[68] American Hospital Association 2024 *America's Hospitals and Health Systems Continue to Face Escalating Operational Costs and Economic Pressures as They Care for Patients and Communities.*
[69] Perdikis S and Millán J d R 2020 Brain-machine interfaces: a tale of two learners *IEEE Sys. Man Cybern. Mag.* **6** 12–19
[70] Perdikis S, Tonin L, Saeedi S, Schneider C and Millán J del R 2018 The Cybathlon BCI race: successful longitudinal mutual learning with two tetraplegic users *PLOS Biol.* **16** e2003787
[71] Collinger J L, Wodlinger B, Downey J E, Wang W, Tyler-Kabara E C, Weber D J, McMorland A J, Velliste M, Boninger M L and Schwartz A B 2013 High-performance neuroprosthetic control by an individual with tetraplegia *Lancet* **381** 557–64
[72] Sauseng P, Klimesch W, Gerloff C and Hummel F C 2009 Spontaneous locally restricted EEG alpha activity determines cortical excitability in the motor cortex *Neuropsychologia* **47** 284–8
[73] Cervera M A, Soekadar S R, Ushiba J, Millán J d R, Liu M, Birbaumer N and Garipelli G 2018 Brain-computer interfaces for post-stroke motor rehabilitation: a meta-analysis *An.n Clin. Transl. Neurol.* **5** 651–63
[74] Biasiucci A *et al* 2018 Brain-actuated functional electrical stimulation elicits lasting arm motor recovery after stroke *Nat. Commun.* **9** 2421
[75] Ruffino C, Gaveau J'emie, Papaxanthis C and Lebon F 2019 An acute session of motor imagery training induces use-dependent plasticity *Sci. Rep.* **9** 20002
[76] Riashad Foysal K M and Baker S N 2020 Induction of plasticity in the human motor system by motor imagery and transcranial magnetic stimulation *J. Physiol.* **598** 2385–96
[77] Di Rienzo F, Blache Y, Kanthack T F, Monteil K, Collet C and Guillot A 2015 Short-term effects of integrated motor imagery practice on muscle activation and force performance *Neuroscience* **305** 146–56
[78] Mizuguchi N and Kanosue K 2017 Changes in brain activity during action observation and motor imagery: their relationship with motor learning *Prog. Brain Res.* **234** 189–204
[79] Tariq M, Trivailo P M and Simic M 2018 Eeg-based bci control schemes for lower-limb assistive-robots *Front. Hum. Neurosci.* **12** 312
[80] Kumar S *et al* 2023 Cognidavinci: towards estimating mental workload modulated by visual delays during telerobotic surgery-an EEG-based analysis *2023 IEEE Int. Conf. on Robotics and Automation (ICRA)* (IEEE) pp 6789–94
[81] Hu Liu D and Anas Imtiaz S 2020 Studying the effects of compression in EEG-based wearable sleep monitoring systems *IEEE Access* **8** 168486–501
[82] Kaposzta Z, Czoch A, Mukli P, Stylianou O, Hu Liu D, Eke A and Samuel Racz F 2024 Fingerprints of decreased cognitive performance on fractal connectivity dynamics in healthy aging *GeroScience* **46** 713–36
[83] Attila Horvath A *et al* 2021 Subclinical epileptiform activity accelerates the progression of Alzheimer's disease: a long-term eeg study *Clin. Neurophys.* **132** 1982–9

[84] Cassani R et al 2018 Systematic review on resting-state EEG for Alzheimer's disease diagnosis and progression assessment *Dis. Markers* **2018** 5174815
[85] Leocani L, Rovaris M, Martinelli Boneschi F, Medaglini S, Rossi P, Martinelli V, Amadio S and Comi G 2006 Multimodal evoked potentials to assess the evolution of multiple sclerosis: a longitudinal study *J. Neurol. Neurosurgery Psychiatry* **77** 1030–5
[86] Taylor M J, Chan-Lui W Y and Logan W J 1985 Longitudinal evoked potential studies in hereditary ataxias *Canadian J. Neurological Sci.* **12** 100–5
[87] Ding Q et al 2024 Neurophysiological characterization of stroke recovery: a longitudinal TMS and EEG study *CNS Neurosci. Ther.* **30** e14471
[88] Fleury L et al 2022 Toward individualized medicine in stroke–the TiMeS project: protocol of longitudinal, multi-modal, multi-domain study in stroke *Frontiers Neurol.* **13** 939640
[89] von Groll V G, Leeuwis N, Rimbert S, Roc A, Pillette L, Lotte F and Alimardani M 2024 Large scale investigation of the effect of gender on mu rhythm suppression in motor imagery brain-computer interfaces *Brain Comput. Int.* 11 87–97
[90] Fischer A G, Danielmeier C, Villringer A, Klein T A and Ullsperger M 2016 Gender influences on brain responses to errors and post-error adjustments *Sci. Rep.* **6** 24435
[91] Liu Q, Yang L, Zhang Z, Yang H, Zhang Y and Jinglong W 2023 The feature, performance and prospect of advanced electrodes for electroencephalogram *Biosensors* **13** 101
[92] Peng H-L, Liu J-Q, Tian H-C, Dong Y-Z, Yang B, Chen X and Yang C-S 2016 A novel passive electrode based on porous ti for eeg recording *Sens. Actuators* B **226** 349–56
[93] Gao K-P, Yang H-J, Liao L-L, Jiang C-P, Zhao N, Wang X-L, Xiu-Yan Li, Chen X, Yang B and Liu J 2019 A novel bristle-shaped semi-dry electrode with low contact impedance and ease of use features for eeg signal measurements *IEEE Trans. Biomed. Eng.* **67** 750–61
[94] Pedrosa P, Fiedler P, Schinaia L, Vasconcelos B, Martins A C, Amaral M H, Comani S, Haueisen J and Fonseca C 2017 Alginate-based hydrogels as an alternative to electrolytic gels for rapid eeg monitoring and easy cleaning procedures *Sens. Actuators* B **247** 273–83
[95] Xing X, Pei W, Wang Y, Guo X, Zhang H, Xie Y, Gui Q, Wang F and Chen H 2018 Assessing a novel micro-seepage electrode with flexible and elastic tips for wearable eeg acquisition *Sens. Actuators A: Phys.* **270** 262–70
[96] Guangli G, Zhang D, Wang S and Duan Y Y 2016 Novel passive ceramic based semi-dry electrodes for recording electroencephalography signals from the hairy scalp *Sens. Actuators* B **237** 167–78
[97] Wang F, Guangli Li, Chen J, Duan Y and Zhang D 2016 Novel semi-dry electrodes for brain–computer interface applications *J. Neural Eng.* **13** 046021
[98] Lin C-T, Lun-De L, Yu-Hang Liu I-J W, Lin B-S and Chang J-Y 2010 Novel dry polymer foam electrodes for long-term EEG measurement *IEEE Trans. Biomed. Eng.* **58** 1200–7
[99] Hongbian Li et al 2023 Hair-compatible sponge electrodes integrated on VR headset for electroencephalography *Soft Sci.* **3** 22