PAPER

Applying dynamic data collection to improve dry electrode system performance for a P300-based brain–computer interface

To cite this article: J M Clements *et al* 2016 *J. Neural Eng.* **13** 066018

View the article online for updates and enhancements.

Applying dynamic data collection to improve dry electrode system performance for a P300-based brain–computer interface

J M Clements[1]**, E W Sellers**[2]**, D B Ryan**[2]**, K Caves**[1]**, L M Collins**[1] **and C S Throckmorton**[1]

[1] Duke University, Durham, 27708, USA
[2] East Tennessee State University, Johnson City, 37604, USA

E-mail: jillian.clements@duke.edu

CrossMark

Abstract

Objective. Dry electrodes have an advantage over gel-based 'wet' electrodes by providing quicker set-up time for electroencephalography recording; however, the potentially poorer contact can result in noisier recordings. We examine the impact that this may have on brain–computer interface communication and potential approaches for mitigation. *Approach.* We present a performance comparison of wet and dry electrodes for use with the P300 speller system in both healthy participants and participants with communication disabilities (ALS and PLS), and investigate the potential for a data-driven dynamic data collection algorithm to compensate for the lower signal-to-noise ratio (SNR) in dry systems. *Main results.* Performance results from sixteen healthy participants obtained in the standard static data collection environment demonstrate a substantial loss in accuracy with the dry system. Using a dynamic stopping algorithm, performance may have been improved by collecting more data in the dry system for ten healthy participants and eight participants with communication disabilities; however, the algorithm did not fully compensate for the lower SNR of the dry system. An analysis of the wet and dry system recordings revealed that delta and theta frequency band power (0.1–4 Hz and 4–8 Hz, respectively) are consistently higher in dry system recordings across participants, indicating that transient and drift artifacts may be an issue for dry systems. *Significance.* Using dry electrodes is desirable for reduced set-up time; however, this study demonstrates that online performance is significantly poorer than for wet electrodes for users with and without disabilities. We test a new application of dynamic stopping algorithms to compensate for poorer SNR. Dynamic stopping improved dry system performance; however, further signal processing efforts are likely necessary for full mitigation.

Keywords: brain–computer interface, electroencephalography, dry electrodes, artefact, dynamic data collection, P300 speller

(Some figures may appear in colour only in the online journal)

1. Introduction

A brain–computer interface (BCI) system allows a person with severe motor impairments to control a device by interpreting measurements of electrical activity in the brain. While head and eye tracking systems may provide alternative methods for controlling a device, these systems still require

some muscle function and can fail when this muscle function is lost (e.g., [1]). Thus, BCIs may provide communication assistance for people with disabilities that limit the utility of other communication aids due to their reliance on physical motion. BCIs that rely on various components of measured electroencephalographic (EEG) data have been proposed in the literature (see [2] for a review); one of which that has

shown potential as a communication device for people with severe physical limitations is the P300 speller [3]. Operation of the P300 speller relies on eliciting event-related potentials (ERPs) to enable the user to spell text on a computer letter by letter. An ERP is a neural response observed in human brain activity, and is produced by the brain's reaction to a specific task or stimulus. The P300 component of the ERP is a positive voltage peak that occurs around 300 ms after the presentation of a rare but relevant stimulus. The name P300 speller refers to the reliance of the speller on this evoked component in the EEG to determine the BCI user's intent, although other ERPs may also be used for classification [4].

Despite encouraging improvements over the last decade, the P300 speller remains primarily a research device rather than a home-based communication aid. One limitation that inhibits its widespread use is the complex and time-consuming process of applying the electrodes to the scalp prior to EEG recording. Although an EEG-based approach has the advantage over implanted electrodes because it is immediately accessible to all users since it does not require surgery, it produces poorer quality signals because neural activity must be measured after it traverses layers of the skull, skin, and hair to reach the electrodes. In order to reduce skin impedance and ensure electrical contact with the scalp, gel-based 'wet' electrodes require abrasive skin preparation and the application of conductive gel to each electrode. This process can require 10 min or more for an experienced lab technician. For people with severe communication disabilities who need to use the system on a daily basis, minimizing this set-up time is highly desirable [5].

In an attempt to reduce the required set-up time, the use of dry electrodes for P300-based BCIs has been explored [6–9]. Dry electrodes are much faster and easier to apply, as they do not require the application of gel to the scalp. Instead, contact dry electrodes contain several metal pins with sufficient length to reach through hair to the skin, coupling signals both resistively and capacitively. While noncontact dry electrodes, which operate primarily via capacitive coupling, have been shown to successfully record steady-state visual evoked potentials, signal quality is reduced when compared to contact dry electrodes [10]. Therefore, the focus of this paper will be on contact dry electrodes. Active dry electrodes employ a local high input impedance amplifier attached to each electrode that rejects interference at the recording site. However, due to poor skin-to-electrode contact, dry systems are much more sensitive to noise and ERP variation [11]. The lower signal-to-noise ratio (SNR) can decrease the detectability of ERPs used for BCI control, which may negatively impact both the speed and accuracy of the system. For this reason, wet electrodes remain the standard for P300-based BCIs.

Several studies have evaluated the performance of dry electrodes in comparison to wet electrodes with P300-based BCI technology in healthy participants with mixed results. In a comparison of the hybrid dry electrode sensor array system to a conventional wet system, Sellers *et al* [6] observed higher P300 speller accuracy with the wet system for four out of eight healthy participants, although the difference in accuracies was not statistically significant. Zander *et al* [7] introduced a dry system developed by Brain Products that only used three electrodes, but offline analysis revealed that a statistically significant difference existed between the P300 classification accuracies of the dry and wet system recordings. Grozea *et al* [8] introduced a novel dry system using bristle sensors, but observed that statistically significant differences exist between the wet and dry P300 signals for each participant. A study performed by Guger *et al* [9] compared P300 speller accuracies with the g.Sahara dry system to accuracies obtained in a previous study [12] with a conventional wet system. The subject pools were different for these two studies. The wet system study [12] reported an average accuracy of $91.0\% \pm 18.5\%$ across 81 subjects while the average accuracy in the dry system study [9] was $90.4\% \pm 17.2\%$ across 23 subjects. Although no significant difference was observed between the accuracies across subject pools for the two systems, intra-subject variation across the two subject pools may have been a confounding factor.

The P300 speller typically collects multiple repetitions of the EEG response and averages them to improve the SNR of the P300 component. In the standard data collection environment, the number of repetitions collected for the average is held constant across participants and target characters (termed 'static' data collection). All of the previous 'wet versus dry' studies mentioned [6–9] relied on static data collection for which the amount of data collected was identical for the wet and dry systems. However, given the poorer quality of the dry electrode recordings, these systems may have benefited from additional data collection to improve the SNR. In this study, we assess the potential for a data-driven dynamic stopping algorithm to compensate for the lower SNR of a dry system by automatically determining the amount of data to collect online in both healthy participants and participants with amyotrophic lateral sclerosis (ALS). Several methods for adaptively controlling data collection have been proposed in literature (e.g., [13–15]). However, these methods typically depend on some form of past performance of the subjects. For example, Serby *et al* [13] and Lendhardt *et al* [14] set a threshold for stopping data collection using averaged training data across a subject pool, which may not translate if the subject pool is changed. Throckmorton *et al* [16] presented a probabilistic data collection algorithm that, based on the quality of in-coming data, automatically determined the amount of data to collect on a character-by-character and subject-by-subject basis. If data were of poor quality, additional data was collected to improve accuracy, and if data were of high quality, data collection was reduced to increase spelling speed. This algorithm resulted in significant improvements in accuracy and spelling speed for healthy participants. These results were validated in a study with participants with ALS, one of the target populations for P300 spellers [17].

The hypothesis for this study is that the dynamic stopping algorithm will detect the need for additional data collection with the dry system and collect sufficient data to improve the BCI performance to wet system levels. Three experiments are conducted. In the first experiment, P300

Table 1. Demographic information of study participants with communication difficulties.

| Participant number	ALSFRS-R	Experiment location	Additional information
1	4	Clinic	Movement via wheelchair, respirator-dependent, feeding tube
2	25	Home	Movement via wheelchair
3	20	Home	Movement via wheelchair
4	20	Home	Ambulatory
5	46	Home	Ambulatory
6	8	Home	Movement via wheelchair, respirator-dependent, feeding tube
7	39	Home	Ambulatory
8	41	Clinic	Ambulatory

speller performances with the g.Sahara dry system and a conventional wet system are compared using static data collection for healthy participants in order to assess differences in performance for consistent amounts of collected data across systems.

The second experiment applies the dynamic stopping algorithm [16] to online data collection and the P300 speller performances for healthy participants using wet and dry systems are compared. This will assess the algorithm's ability to automatically compensate for any potential SNR differences between the two systems by increasing data collection as needed. Finally, results with the dynamic stopping algorithm are evaluated in participants with ALS. An analysis of the differences in signal characteristics between EEG collected with wet and dry systems was conducted to clarify the observed performance results.

2. Methods

2.1. Participants and equipment

In the first experiment of this study, P300 speller data was collected in the standard static data collection environment for 16 healthy participants at East Tennessee State University. The second experiment used the dynamic stopping algorithm [16] for online data collection for ten healthy participants at Duke University. The third experiment recruited eleven participants with communication disabilities (ALS = 10, PLS = 1) to validate the results observed for the healthy participants in the second experiment. Three participants with ALS were dropped from the study due to vision loss or caregiver unavailability. Details of the study participants with communication disabilities used for data analysis can be found in table 1. Participants in each experiment completed two P300 speller sessions; one using a wet system and one using a dry system. The two sessions were performed between 1 and 54 d apart for the healthy participants and between 1 and 95 d apart for the participants with communication disabilities. The electrode type used in the first session was counter-balanced across participants. Participants were asked to remove any HF-communication devices (e.g., mobile phones) that may interfere with the system and to avoid any type of movement during both sessions. This is imperative for dry electrode recordings, which may be more
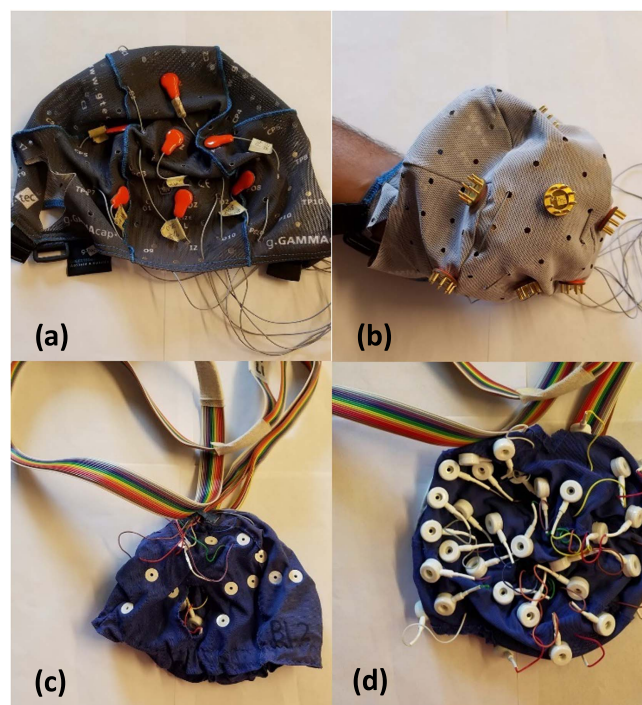
Figure 1. This figure shows the caps used for this study. The g. Sahara dry electrode cap is shown for the (a) outside and (b) inside of the cap. The Electro-cap International wet electrode cap is shown for the (c) outside and (d) inside of the cap.

sensitive to these types of interferences. P300 speller sessions with healthy participants were performed in a laboratory setting, while sessions with participants with disabilities were performed in the Duke ALS clinic and participants' homes. All experiments were approved by either the East Tennessee State University or Duke University IRB boards.

EEG signals were measured using electrodes positioned according to the International 10–20 system and connected to a computer via a 16-channel g.USBAMP biosignal amplifier. The signals were sampled at 256 Hz. The dry electrode cap utilized the g.Sahara active dry system, which is comprised of 8-pin golden alloy coated electrodes (7 mm pin length). The wet electrode cap and electrolyte solution were purchased from Electro-Cap International. A picture of each cap is shown in figure 1. The dry electrode cap includes a chin strap and the wet electrode cap does not. Three cap sizes (small, medium, and large) were available for both electrode types.

Figure 2. Screen captures of the character grid used for healthy participants (left) and participants with ALS (right).

Head circumference and the distance from the nasion to inion were measured to determine an appropriate cap size that provides gentle pressure for each participant. The caps were pulled over the ears and were then connected to the amplifier. Grounding bracelets were worn by both the experimenter and the participant during dry electrode recording sessions to avoid electrostatic charges. Eight electrodes (Fz, Cz, P3, Pz, P4, P07, P08, and Oz), with ground and reference electrodes attached to each mastoid, were used for data collection and classification. Dry electrodes were turned back and forth a few times to remove any hair between the electrode and the skin, ensuring proper electrical contact. These electrodes have been demonstrated to provide adequate information for P300 speller communication [18]. The BCI2000 software package [19] was used for stimulus presentation and data collection, with additional functionality added for the application of the dynamic stopping algorithm [16].

2.2. P300 speller paradigm

Healthy participants were presented with a 9×8 grid of characters on a computer screen while participants with communication disabilities were presented with a 6×6 grid of characters (see figure 2). These grids match the grids used in Throckmorton *et al* [16] and Mainsah *et al* [17] for participants without disability and with ALS, respectively. Both grids were flashed based on the checkerboard paradigm to eliminate character adjacency and double flash errors [20]. Each character stimulus was flashed twice in a sequence of 24 flashes. The flash duration was 62.5 ms followed by an inter-stimulus interval of 62.5 ms, with an inter-target interval of 3.5 s. EEG signals were recorded for 800 ms following each flash. The number of EEG signals collected for a given target character differ between experiments and are detailed in the following subsections.
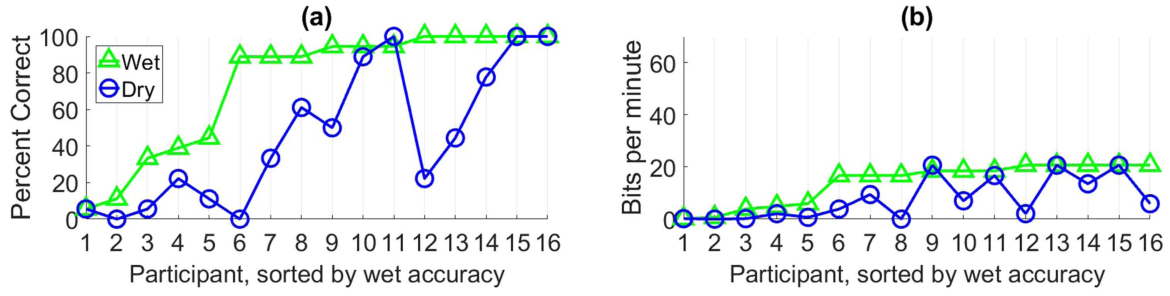
2.3. Static data collection

Three calibration runs and three online test runs were recorded for both the wet and dry system sessions in the static data collection experiment. For each run, the participant was asked

to copy-spell a six-character token randomly drawn from a subset of tokens from the English language. EEG response signals to five sequences or 120 flashes (5 sequences $\times$ 24 flashes/sequence = 120 flashes) were collected for each target character presented to the participant. These data were preprocessed and features were extracted for classification according to Krusienski *et al* [18]. Stepwise linear discriminate analysis (SWLDA) was performed on the calibration data to create the classifier weights for online testing. Although several learning algorithms are known to adequately classify P300 speller data, a statistical analysis of several classifiers has indicated that the SWLDA classifier is sufficient and the added complexity of nonlinear methods is not necessary [21].

2.4. Dynamic data collection

Calibration runs for the dynamic data collection experiment were gathered in a similar manner to static data collection. However, due to the performance results of the static data collection experiment, five calibration runs were collected instead of three to improve the weights of the classifier. For each calibration run, the participant was asked to copy-spell a six-character token from the English language. EEG response signals to seven sequences or 168 flashes (7 sequences $\times$ 24 flashes/sequence = 168 flashes) were collected for each character. The preprocessing techniques, feature extraction and classification method were identical to the static experiment.

Five online test runs were collected with the dynamic stopping algorithm [16] applied. Instead of having a pre-set number of flash sequences for data collection, the dynamic stopping algorithm automatically determined the necessary amount of data to collect for each target character. The amount of data collected was controlled by a threshold (90%) on the probability that each character in the grid was the target character (c^*). The initial probability of each character being the target was set to $1/N$ (where N is the total number of characters in the grid), i.e., no *a priori* knowledge was assumed. For each new response signal (x_j) obtained after a character group flash (S_j contains the characters in the group),

Figure 3. This figure shows the results for the static data collection experiment performance comparison of wet (triangles) and dry (circles) electrode systems for sixteen healthy participants. The *x*-axis represents the participants in this study and the *y*-axis represents the results for (a) spelling accuracy, in percent correct, and (b) bit rate, in bits min^{-1}.

a classifier score ($y_j = w * x_j$, where w represents the classifier weights) was calculated and stored in $Y_j = [y_1, \ldots, y_{j-1}, y_j]$. The probability for each character (c_n) in group S_j was updated using the likelihood estimates from the training data (obtained from the probability density functions for the target, $p(y_i|H_1)$, and non-target, $p(y_i|H_0)$, response signals). The update equation is based on Bayes rule and implemented as in equation (1), where $P(c_n = c^*|Y_j, S_j)$ is the posterior probability of a character c_n being the target character, $P(y_j|c_i = c^*, S_j)$ is the likelihood of the classifier response given that c_i was/was not in group S_j and $P(c_i = c^*|Y_{j-1}, S_{j-1})$ is the prior probability of a character c_i being the target character.

$$P(c_n = c^*|Y_j, S_j)$$
$$= \frac{P(y_j | c_n = c^*, S_j)P(c_n = c^* | Y_{j-1}, S_{j-1})}{\sum_{i=1}^{N} P(y_j | c_i = c^*, S_j)P(c_i = c^* | Y_{j-1}, S_{j-1})}. \qquad (1)$$

Character probabilities were updated after each response to a flash group was collected. Data collection stopped once one of the character probabilities in the grid exceeded 90%.

3. Results

Figure 3 plots the spelling accuracy and bit rate (related to spelling speed), respectively, for each participant using the dry system (circle) and wet system (triangle) in the static data collection experiment. Accuracy is defined as the percentage of characters spelled correctly in a copy-spelling session. Assuming a binomial distribution on the probability of correctly selecting a set of characters [22], chance level accuracy was 11.1% for this experiment. Bit rate is a measure of communication systems that incorporates accuracy, speed, and the number of selectable characters presented and is calculated in bits min^{-1} [23].

The results in figure 3 illustrate the impact of the lower SNR resulting from the dry system. Twelve out of the sixteen participants performed worse with the dry system when compared to the wet system. The average spelling accuracy and bit rate for the wet and dry systems are listed in table 2. Two Wilcoxon signed-rank tests were performed to compare the participants' accuracy and bit rate with dry and wet systems and the differences were statistically significant ($p = 0.0007$ and $p = 0.0006$, respectively). This implies that

Table 2. Static data collection performance averages.

	Average accuracy (%)	Average bit rate (bits min^{-1})
Dry system	42.8	7.3
Wet system	72.8	13.6

dry systems may benefit from additional data collection to improve the SNR.

The results for the dynamic data collection experiment are shown in figure 4 and the average spelling accuracy and bit rate for both systems are listed in table 3. Chance level accuracy for this experiment was 10.0%. With dynamic data collection, the differences in accuracy between wet and dry systems were small (<13%) for the majority of subjects (see figure 4(a)). However, the differences in accuracy and bit rate were still statistically significant ($p = 0.025$ and $p = 0.002$, respectively) using Wilcoxon signed-rank tests. Although the average performance with dry electrodes remains lower than performance with wet electrodes despite using dynamic collection, the differences in BCI performance between wet and dry systems appear smaller than those observed using static data collection. In order to compare the static results to the dynamic results across the two subject pools, a subset of participants were selected from the static data collection experiment based on their wet system performance. Eleven participants in the static data collection experiment had wet system spelling accuracies above 80%, similar to the ten participants in the dynamic data collection study. The average accuracy for this subset of participants with the wet system was 95.5%, similar to the 94.3% average accuracy observed for the dynamic data collection participants. However, the average dry system accuracy for these participants was 61.6%; still much lower than the 79% average accuracy observed in the dynamic data collection experiment, suggesting that dynamic data collection may mitigate some of the lost performance with dry electrodes.

As hypothesized, the dynamic stopping algorithm increased the amount of data collected for the dry system across all subjects (see figure 4(c)), indicating that the algorithm can detect and respond to the need for additional data collection in low SNR scenarios. However, collecting more
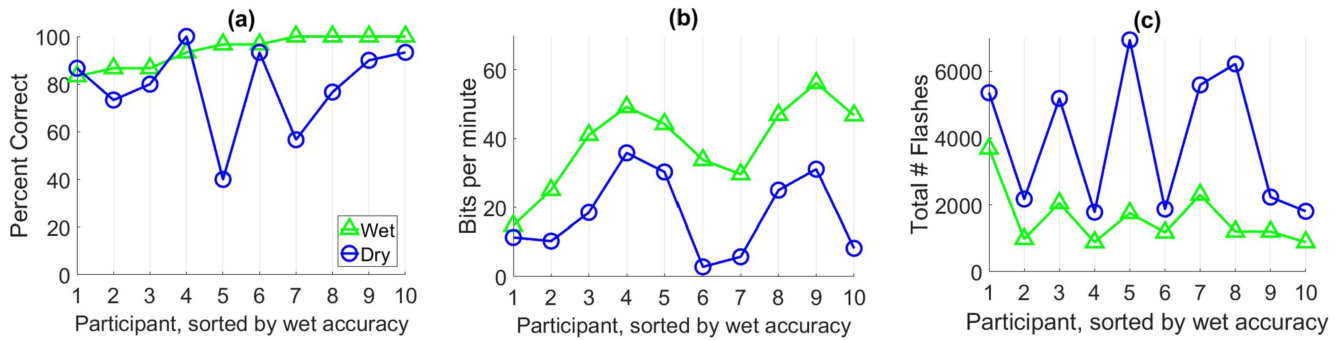
Figure 4. This figure shows the results for the dynamic data collection experiment performance comparison of wet (triangles) and dry (circles) electrode systems for healthy participants. The *x*-axis represents the participants in this study and the *y*-axis represents the results for (a) spelling accuracy, in percent correct, and (b) bit rate, in bits min^{-1}, and (c) total number of flashes collected.

Table 3. Dynamic data collection (healthy participants) performance averages.

|	Average accuracy (%)	Average bit rate (bits min^{-1})	Average # flashes
Dry system	79.0	17.9	3915
Wet system	94.3	38.7	1615

data did not increase the performance of the dry systems to the same level of performance as the wet systems.

The question of whether the raw data from the two electrode types looked similar was also addressed. In figure 5, P300 latency and amplitude values were compared for signals measured over the electrode location Cz in the dynamic data collection study. These values were selected by taking the maximum from a window of the grand average between 0.2–0.4 s. The grand averages include P300 signals from both the training and testing stages; therefore, more data were averaged for the dry grand averages than the wet grand averages (see figure 4(c)). The average P300 latency for the wet and dry systems were 0.262 s ± 0.031 and 0.263 s ± 0.032, respectively, and the differences between systems were not statistically significant ($p = 0.688$). However, the differences in P300 amplitude were statistically significant ($p = 0.027$). The average dry system amplitude (2.55

μV ± 1.14) was lower than the wet system amplitude (3.29 μV ± 1.65), which may indicate that techniques other than additional data collection may be needed to improve the SNR of the dry system.

The results for healthy participants in the dynamic data collection experiment were confirmed for participants with communication disabilities (see figure 6 and table 4). Once again, the differences in accuracy and bit rate were statistically significant ($p = 0.008$ and $p = 0.008$, respectively) and the amount of data collected increased across all participants for the dry system. This further implies that the dynamic stopping algorithm improves the SNR of the dry system by collecting more data, but performance is still degraded when compared to the wet system. Improvements, either in signal processing or in hardware design, may be required to bring dry systems to the same level of performance as wet systems.

Participants with communication disabilities were also given a short survey after both the wet and dry system sessions. The survey consisted of six statements and the participants were asked to rate each statement with a number between one (strongly disagree) and seven (strongly agree). The average responses are shown in table 5. These averages suggest that while the participants agree that the setup of the dry system was quick and not agitating, they also found the dry system to be slow and uncomfortable. However, the
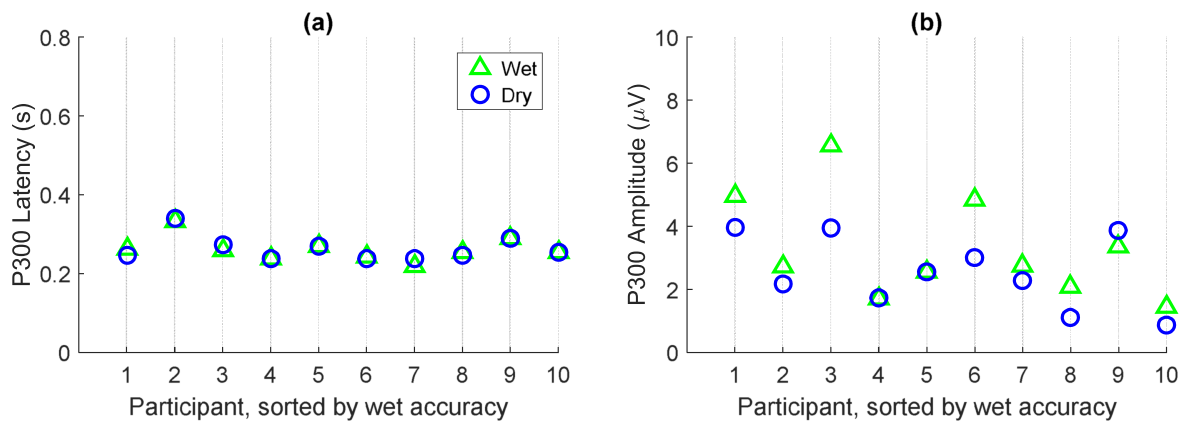
Figure 5. This figure shows the latency and amplitude values for the grand average P300 signals over electrode location Cz in the dynamic data collection experiment. The *x*-axis represents the healthy participants in this study and the *y*-axis represents the P300 values for (a) latency, in seconds, and (b) amplitude, in microvolts.
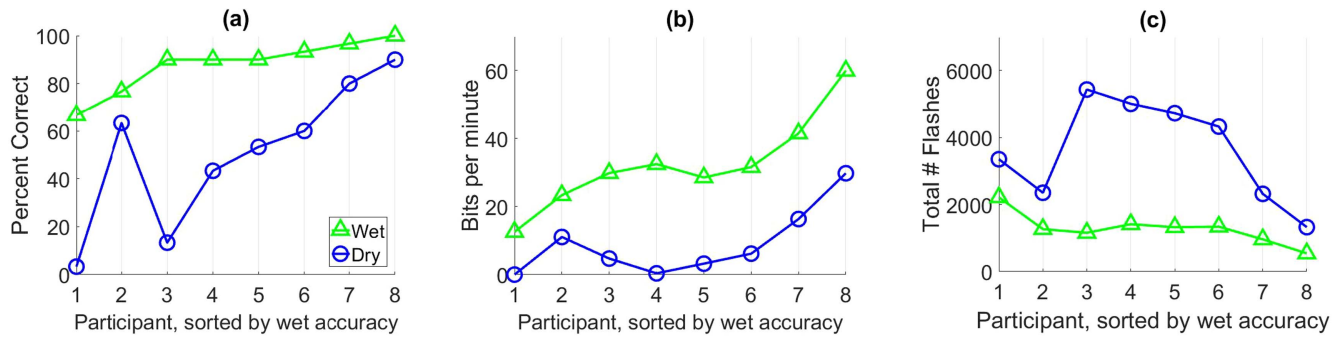
Figure 6. This figure shows the results for the dynamic data collection experiment performance comparison of wet (triangles) and dry (circles) electrode systems for participants with communication disabilities. The *x*-axis represents the participants in this study and the *y*-axis represents the results for (a) spelling accuracy, in percent correct, and (b) bit rate, in bits per minute, and (c) total number of flashes collected.

Table 4. Dynamic data collection (ALS/PLS participants) performance averages.

	Average accuracy (%)	Average bit rate (bits min^{-1})	Average # flashes
Dry system	50.9	8.9	3604
Wet system	87.9	32.5	1278

differences between the wet and dry system responses for each statement were not statistically significant.

In order to determine the types of signal processing most likely to be beneficial for improving the performance of dry systems, the difference(s) between wet and dry electrode recordings need to be characterized. In a wet system, the conductive gel serves as buffer to bridge the gaps between the skin and the electrode as a participant moves during a recording session. Without this conductive gel, dry systems may be more sensitive to artefacts from movement. Therefore, one potential difference between data collected with the two systems may be the number of motion artefacts recorded.

Motion artefacts create large potentials that are superimposed onto neural signals and are known to impact the frequency content of EEG recordings [24]. These artefacts can have amplitudes of several millivolts and often have energy in the same frequency range as EEG events, such as the P300 component. To determine whether an increase in spectral amplitude occurred for the dry system in our experiments, the power spectral density (PSD) estimates were examined for each participant's continuous dry and wet electrode recordings. Figure 7 displays the dry and wet PSD results for a single electrode channel, Cz, in the dynamic data

collection experiments with healthy participants (figures 7(a) and (b)) and participants with communication disabilities (figures 7(c) and (d)). Estimates below 3 dB and above 30 dB were clipped for visualization purposes. The PSD results displayed in figure 7 indicate that low-frequency noise may be higher in dry electrode recordings when compared to wet electrode recordings.

The difference in frequency content between data collected with the wet and dry systems was further analyzed by examining four frequency band powers: (1) delta (0.1–4 Hz), (2) theta (4–8 Hz), (3) alpha (8–13 Hz), and (4) beta (13–30 Hz). Delta and theta band power were repeatedly higher in the continuous dry electrode recordings than wet electrode recordings. A statistical analysis of the wet and dry power bands across all participants, using Wilcoxon signed-rank tests, demonstrated that these differences in delta and theta band power between the two systems were significant for 8 out of the 8 electrode channels ($p < 0.05$). The alpha band power differences were significant for seven channels, while the differences for the beta band power were significant for three channels. This indicates that as frequency decreases, the differences between wet and dry system recordings increases across channels. A spectral analysis of all eight electrode channels demonstrated that the low-frequency power is generally higher for the dry system recordings, which suggests that low-frequency artefact noise may be negatively impacting dry system performance given that the P300 component also lies in this frequency band [25–27].

To measure possible artefact noise, two techniques that have been well-studied in seizure detection literature (e.g., [28–33]) were used to detect large transient voltages in P300 speller EEG recordings between characters (when characters

Table 5. Survey results for participants with communication difficulties.

| Statement (1 = Strongly disagree, 7 = Strongly agree)	Wet (Average)	Dry (Average)
(1) The time required to setup the cap was reasonable	5.70 (±2.20)	6.50 (±0.79)
(2) The setup of the cap was frustrating/agitating	2.28 (±1.89)	1.00 (±0.00)
(3) I found the cap comfortable	5.42 (±1.81)	4.00 (±1.95)
(4) I could see myself wear this cap for more than two hours at a time	5.00 (±2.08)	2.50 (±2.07)
(5) The speed and accuracy of the system was acceptable	5.71 (±1.11)	3.25 (±2.70)
(6) If this is my primary method of controlling a computer I would use it daily	5.14 (±2.27)	2.75 (±1.98)
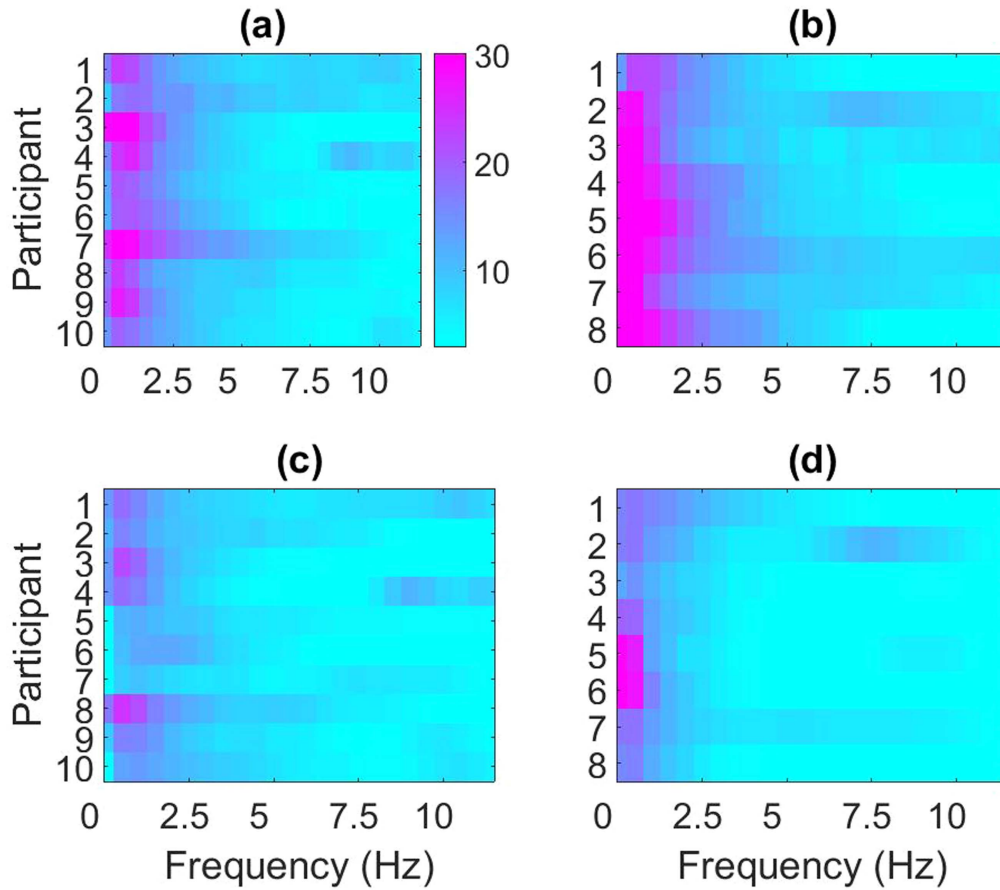
Figure 7. This figure displays the power spectral density estimates for continuous P300 speller EEG data. The estimates are shown for ten healthy participants using the (a) dry system and (b) wet system and eight participants with communication disabilities using the (c) dry system and (d) wet system. The *y*-axis denotes the participant and the *x*-axis denotes frequency (in Hz). The color axis displays the power (in dB), where the values are mapped to a color according to the color bar on the right (e.g., power values $\geqslant 30$ dB are shown in solid white and values $\leqslant 3$ dB are shown in solid black).
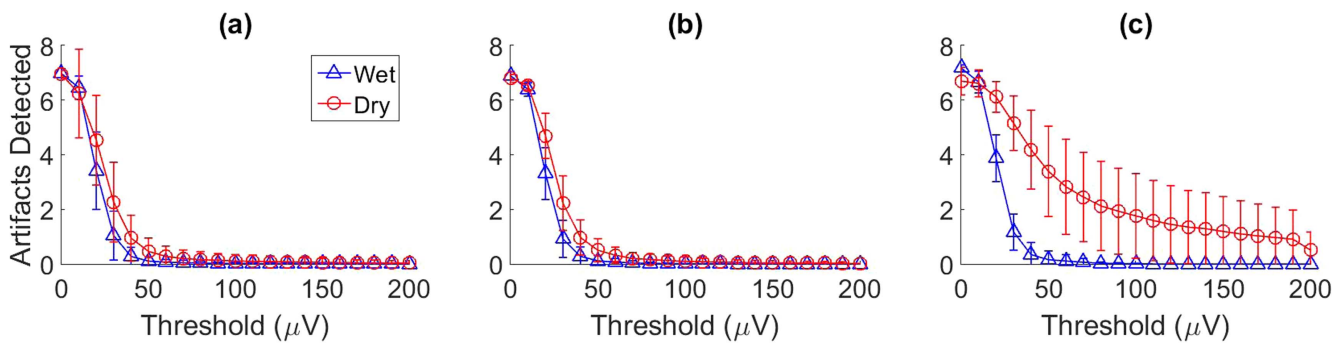
Figure 8. These plots show the results for possible artefact detection using an autoregressive model. The results were averaged across participants in (a) the healthy static data collection experiment, (b) the healthy dynamic data collection experiment and (c) the ALS dynamic data collection experiment. Each plot displays the number of possible artefacts detected (*y*-axis) for a threshold (*x*-axis) in the dry electrode recordings (circles) and the wet electrode recordings (triangles).

were not flashing, eliminating the possibility of inadvertently detecting ERPs). The first technique used an autoregressive model of order 7 on the inter-target intervals of the EEG data (3.5 second segments), where the model coefficients were estimated with the Yule Walker equations. Assuming transient artefacts are non-stationary phenomena, the model detected large peaks in error (exceeding a given threshold) when artefacts may have occurred. The mean number of

artefacts detected for a set of thresholds were averaged across the participants in each experiment and the results are displayed in figure 8. Although the differences between the number of possible artefacts detected with the wet and dry systems for healthy participants seem small (figures 8(a) and (b)), two Wilcoxon sign-ranked test revealed that statistically significant differences exist ($p < 0.001$ for both the static and dynamic experiments). The second technique used wavelet
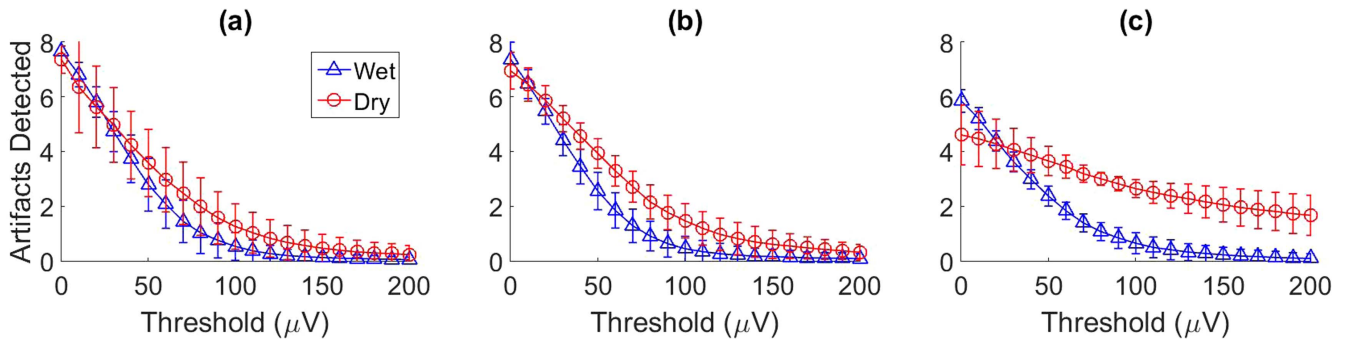
Figure 9. This figure shows the results for possible artifact detection using wavelet analysis. The results were averaged across participants in (a) the healthy static data collection experiment, (b) the healthy dynamic data collection experiment and (c) the ALS dynamic data collection experiment. Each plot displays the number of possible artifacts detected (*y*-axis) for a threshold (*x*-axis) in the dry electrode recordings (circles) and the wet electrode recordings (triangles).

analysis with a Daubechies-4 mother wavelet [31]. Peaks in the low-frequency wavelet approximation coefficients above a certain threshold were measured as artefacts. The results were averaged across the participants in each experiment and are shown in figure 9. Once again, the differences in the number of possible artefacts detected with the wet and dry systems for healthy participants seem small (figures 9(a) and (b)), but are indeed statistically significant ($p < 0.001$ for both the static and dynamic experiments). The results in figures 8 and 9 show that for thresholds above 30–40 μV, both techniques detected a larger number of possible artefacts in the dry electrode recordings when compared to the wet electrode recordings. The results for participants with communication disabilities (figures 8(c) and 9(c)) reveal much larger differences between the number of possible artefacts detected for the wet and dry systems. Given that these experiments were performed either in the participant's home or a clinical setting (as opposed to the healthy participants who performed experiments in a laboratory setting), the larger differences may be due to an increase in environmental noise that dry electrodes, with their high impedance levels, may be more sensitive to such as artefacts from external devices or electrostatic charges. These results suggest that there may be a benefit to applying signal processing approaches that target low-frequency noise without attenuating low-frequency ERPs such as the P300 component.

4. Discussion

While some previous studies have suggested that dry electrode recordings are comparable to wet electrode recordings for P300-based BCI applications, our performance results suggest that dry electrodes produce lower SNRs in the recorded responses, which negatively impacts classification accuracy and reduces their potential utility for a home-based communication aid. Using a data-driven dynamic stopping algorithm compensates for the additional noise by collecting more data, but performance is still reduced when compared to signals recorded with the wet electrodes. Although a variety of commercially available contact dry electrode systems exist (e.g., systems designed by Cognionics, Brain Products,

Emotiv and Wearable Sensing), the results presented in this study would likely translate to these systems since the issue of poor skin-to-electrode contact remains. However, certain design criteria might reduce some of the artefact noise. A wireless dry system that limits cable movement might have been beneficial for ambulatory participants in this study. Flexible dry electrodes that conform to the scalp could have helped provide the right amount of contact pressure for individual sensors, minimizing participant discomfort and possibly reducing motion artefact noise in cases where the cap was ill-fitted. While testing with additional systems would provide validation of the results, without skin abrasion or a conductive gel to create a stable, low impedance path between the skin and the electrode, dry electrodes are likely to be more susceptible to skin potentials and motion artefacts.

Motion artefacts have a wide frequency range, which is maximal at 30 Hz [24]. An analysis of the recorded EEG signal characteristics revealed that the delta and theta frequency band power (0.1 Hz–4 Hz and 4 Hz–8 Hz, respectively) are consistently higher in dry electrode recordings across participants. Two transient artefact detection techniques (autoregressive modeling and wavelet analysis) revealed an increased number of possible skin and motion artefacts for the dry system when compared to the wet system, suggesting that both transient and slow artefacts may be contributing to the lower SNRs in dry systems, potentially resulting in decreased system performance.

Thus a potential future approach to make dry electrodes in P300-based BCI systems feasible as a home-based communication aid may be to develop methods to mitigate the effects of excessive low-frequency noise. While several artefact removal techniques exist, their application in BCI is limited because they may attenuate the desired signal (e.g., linear filtering [34]), require *a priori* knowledge of the desired/artefact signal (e.g., adaptive filtering [34]), not detect small amplitude artefacts (e.g., thresholding [35]) or require that the artefact and desired signal are minimally correlated or maximally independent (e.g., blind source separation [35]). However, it may be possible to use machine learning to separate undesirable artefacts from ERP responses and thereby improve ERP-based BCI performance with dry electrodes. The current standard learning algorithm operates

under the assumption that an increase in energy only occurs for an appropriate stimulus, ignoring voltages that can originate from the user's physical activity or external movements in the surrounding environment. Dry electrodes, with their higher contact impedance, may be particularly sensitive to these types of artefact interference. Therefore, an additional classification stage could be developed that separates detected artefacts into two classes, ERP responses and artefacts. The exploration of features beyond spatiotemporal characteristics, such as statistical moments and PSD estimates, may help increase artefact classification performance. Unsupervised learning techniques could also facilitate further discrimination between different types of artefacts, leading to improved artefact detection methods. The focus of future work will be on incorporating artefact separation into the dynamic stopping algorithm to reduce the system's response to false energy responses and produce a more accurate learning algorithm in dry electrode systems.

5. Conclusion

Using a dry electrode cap with the P300 speller would greatly reduce the complexity and time it takes to set up the system, creating a more practical home-based communication aid for locked-in patients. In the static data collection environment, performance was shown to drop in dry systems, and our results suggest that collecting more data does not fully compensate for the lower SNR when compared to wet systems. Higher low-frequency power and an increased number of large voltage peaks in the dry electrode recordings suggest that transient artefacts may be an issue. To further improve the performance of dry electrodes, a process to mitigate the effects of low-frequency artefact noise may need to be developed.

Acknowledgments

References

[1] Sellers E W, Vaughan T M and Wolpaw J R 2010 A brain-computer interface for long-term independent home use *Amyotrophic Lateral Sclerosis* **11** 449–55

[2] Nicolas-Alonso L F and Gomez-Gil J 2012 Brain computer interfaces, a review *Sensors* **12** 1211–79

[3] Farwell L A and Donchin E 1988 Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials *Electroencephalogr. Clin. Neurophysiol.* **70** 510–23

[4] Allison B Z and Pineda J A 2003 ERPs evoked by different matrix sizes: implications for a brain computer interface (BCI) system *IEEE Trans. Neural Syst. Rehabil. Eng.* **11** 110–3

[5] Huggins J E, Wren P A and Gruis K L 2011 What would brain-computer interface users want? Opinions and priorities of potential users with amyotrophic lateral sclerosis *Amyotrophic Lateral Sclerosis* **12** 318–24

[6] Sellers E W, Turner P, Sarnacki W A, McManus T, Vaughan T M and Matthews R 2009 A novel dry electrode for brain-computer interface *Human-computer interaction. Novel Interaction Methods and Techniques* (Berlin: Springer) pp 623–31

[7] Zander T O, Lehne M, Ihme K, Jatzev S, Correia J, Kothe C, Picht B and Nijboer F 2011 A dry EEG-system for scientific research and brain-computer interfaces *Front. Neurosci.* **5** 1–10

[8] Grozea C, Voinescu C D and Fazli S 2011 Bristle-sensors—low-cost flexible passive dry EEG electrodes for neurofeedback and BCI applications *J. Neural Eng.* **8** 025008

[9] Guger C, Krausz G, Allison B Z and Edlinger G 2012 Comparison of dry and gel based electrodes for p300 brain-computer interfaces *Front. Neurosci.* **6** 1–7

[10] Chi Y M, Wang Y T, Wang Y, Maier C, Jung T P and Cauwenberghs G 2012 Dry and noncontact EEG sensors for mobile brain-computer interfaces *IEEE Trans. Neural Syst. Rehabil. Eng.* **20** 228–35

[11] Khandpur R S 1992 *Handbook of Biomedical Instrumentation* (India: Tata McGraw-Hill Education) pp 39–62

[12] Guger C, Daban S, Sellers E, Holzner C, Krausz G, Carabalona R, Gramatica F and Edlinger G 2009 How many people are able to control a P300-based brain-computer interface (BCI)? *Neurosci. Lett.* **462** 94–8

[13] Serby H, Yom-Tov E and Inbar G F 2005 An improved P300-based brain-computer interface *IEEE Trans. Neural Syst. Rehabil. Eng.* **13** 89–98

[14] Lenhardt A, Kaper M and Ritter H J 2008 An adaptive P300-based online brain-computer interface *IEEE Trans. Neural Syst. Rehabil. Engineering* **16** 121–30

[15] Jin J, Allison B Z, Sellers E W, Brunner C, Horki P, Wang X and Neuper C 2011 An adaptive P300-based control system *J. Neural Eng.* **8** 036006

[16] Throckmorton C S, Colwell K A, Ryan D B, Sellers E W and Collins L M 2013 Bayesian approach to dynamically controlling data collection in P300 spellers *IEEE Trans. Neural Syst. Rehabil. Eng.* **21** 508–17

[17] Mainsah B O, Collins L M, Colwell K A, Sellers E W, Ryan D B, Caves K and Throckmorton C S 2015 Increasing BCI communication rates with dynamic stopping towards more practical use: an ALS study *J. Neural Eng.* **12** 016013

[18] Krusienski D J, Sellers E W, McFarland D J, Vaughan T M and Wolpaw J R 2008 Toward enhanced P300 speller performance *J. Neurosci. Methods* **167** 15–21

[19] Schalk G, McFarland D J, Hinterberger T, Birbaumer N and Wolpaw J R 2004 BCI2000: a general-purpose brain-computer interface (BCI) system *IEEE Trans. Biomed. Eng.* **51** 1034–43

[20] Townsend G, LaPallo B K, Boulay C B, Krusienski D J, Frye G E, Hauser C, Schwartz N E, Vaughan T M, Wolpaw J R and Sellers E W 2010 A novel P300-based brain-computer interface stimulus presentation paradigm: moving beyond rows and columns *Clin. Neurophysiol.* **121** 1109–20

[21] Krusienski D J, Sellers E W, Cabestaing F, Bayoudh S, McFarland D J, Vaughan T M and Wolpaw J R 2006 A comparison of classification techniques for the P300 speller *J. Neural Eng.* **3** 299

[22] Silvoni S, Volpato C, Cavinato M, Marchetti M, Priftis K, Merico A, Tonin P, Koutsikos K, Beverina F and Piccione F 2009 P300-based brain-computer interface communication: evaluation and follow-up in amyotrophic lateral sclerosis *Front. Neurosci.* **3** 1

[23] McFarland D J, Sarnacki W A and Wolpaw J R 2003 Brain-computer interface (BCI) operation: optimizing information transfer rates *Biol. Psychol.* **63** 237–51

[24] Fatourechi M, Bashashati A, Ward R K and Birch G E 2007 EMG and EOG artifacts in brain computer interface systems: a survey *Clin. Neurophysiol.* **118** 480–94

[25] Duncan-Johnson C C and Donchin E 1979 The time constant in P300 recording *Psychophysiology* **16** 53–5

[26] Bougrain L, Saavedra C and Ranta R 2012 Finally, what is the best filter for P300 detection? *TOBI Workshop lll-Tools for Brain-Computer Interaction 2012*

[27] Fabiani M and Donchin E 1995 Encoding processes and memory organization: a model of the von Restorff effect *J. Exp. Psychol.: Learn., Mem., Cognit.* **21** 224

[28] Arakawa K, Fender D H, Harashima H, Miyakawa H and Saitoh Y 1986 Separation of a nonstationary component from the EEG by a nonlinear digital filter *IEEE Trans. Biomed. Eng.* **7** 724–6

[29] Pfurtscheller G and Fischer G 1978 A new approach to spike detection using a combination of inverse and matched filter techniques *Electroencephalogr. Clin. Neurophysiol.* **44** 243–7

[30] Clark I, Biscay R, Echeverría M and Virués T 1995 Multiresolution decomposition of non-stationary EEG signals: a preliminary study *Comput. Biol. Med.* **25** 373–82

[31] Indiradevi K P, Elias E, Sathidevi P S, Nayak S D and Radhakrishnan K 2008 A multi-level wavelet approach for automatic detection of epileptic spikes in the electroencephalogram *Comput. Biol. Med.* **38** 805–16

[32] Latka M, Was Z, Kozik A and West B J 2003 Wavelet analysis of epileptic spikes *Phys. Rev.* E **67** 052902

[33] Senhadji L and Wendling F 2002 Epileptic transient detection: wavelets and time-frequency approaches. *Neurophysiologie Clinique/Clin. Neurophysiol.* **32** 175–92

[34] Sweeney K T, Ward T E and McLoone S F 2012 Artifact removal in physiological signals—practices and possibilities *IEEE Trans. Inf. Technol. Biomed.* **16** 488–500

[35] Delorme A, Sejnowski T and Makeig S 2007 Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis *Neuroimage* **34** 1443–9